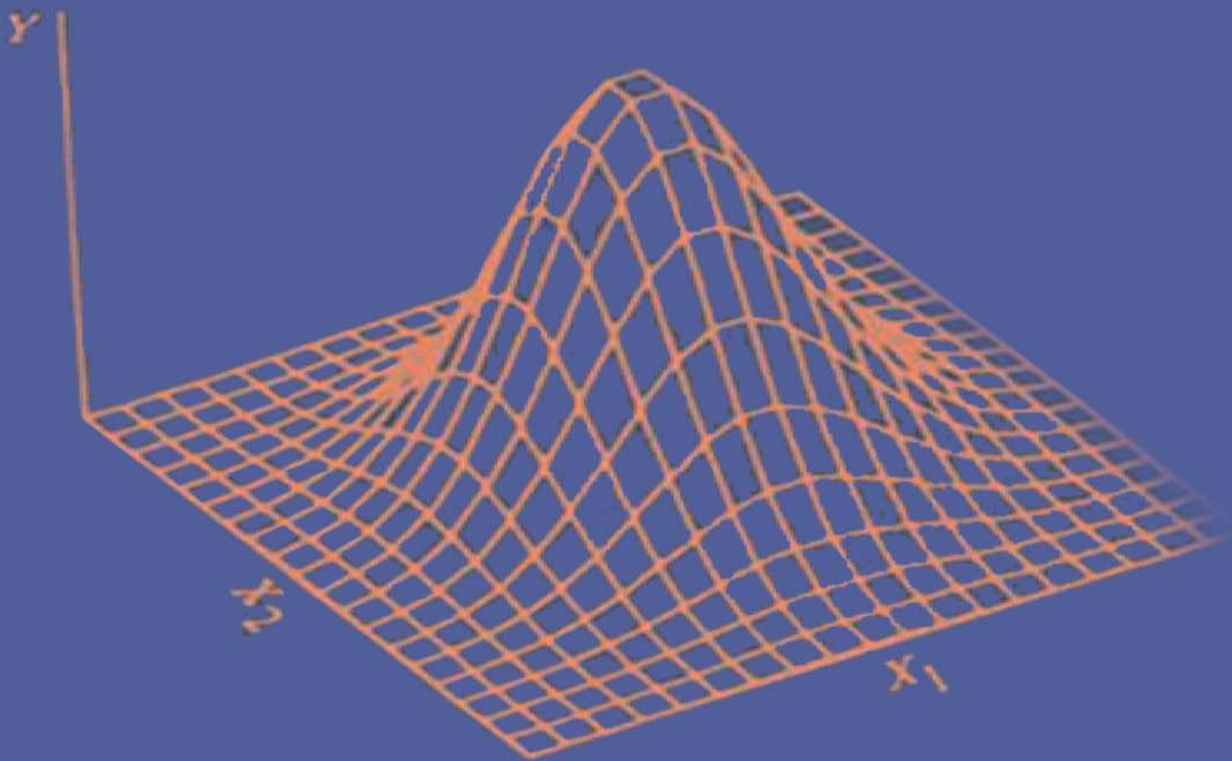
مبادئ الاحصاء

الدكتور
طه حسين الزبيدي





بسم اللـه الرحمن الرحيم

مبادئ الاحصاء



رقم الإيداع لدى المكتبة الوطنية (2012/6/2005)

519.5

الزبيدي، طه حسين

مبادئ الاحصاء/ طه حسين الزبيدي

/ عمان: دار غيداء للنشر والتوزيع، 2012

( ) ص

ر.أ: (2012/6/2005).

الواصفات:/ الاحصاء الوصفي// الاحصاء

❖ تم إعداد بيانات الفهرسة والتصنيف الأولية من قبل دائرة المكتبة الوطنية

ISBN 978-9957-555-51-1



دار غيداء للنشر والتوزيع

تلاع العلي - شارع الملكة رانيا العبدالله | مجمع العساف التجاري - الطابق الأول
تلفاكس : 5353402 6 962+ | خلوي : 95667143 7 962+
ص.ب : 520946 عمان 11152 الأردن | E-mail: darghidaa@gmail.com

مبادئ الاحصاء

الدكتور

طه حسين الزبيدي

كلية الادارة والاقتصاد

جامعة صلاح الدين

الطبعة الأولى

2013م - 1434هـ

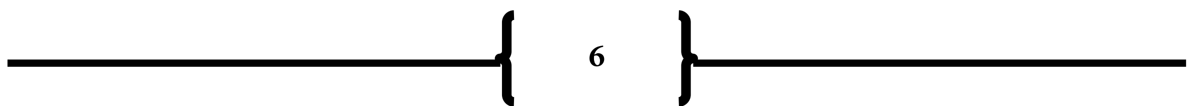


بسم اللـه الرحمن الرحيم

(ذَلِكَ الَّذِي يُبَشِّرُ اللـهُ عِبَادَهُ الَّذِينَ آمَنُوا وَعَمِلُوا الصَّالِحَاتِ قُل لَّا أَسْأَلُكُمْ عَلَيْهِ أَجْرًا إِلَّا الْمَوَدَّةَ فِي الْقُرْبَى وَمَن يَقْتَرِفْ حَسَنَةً نَّزِدْ لَهُ فِيهَا حُسْنًا إِنَّ اللـهَ غَفُورٌ شَكُورٌ)

سورى الشورى الآية (23)

صدق اللـه العظيم



الإهداء

إلى كل من طلب العلم

وإجتهد في تحصيل المعرفة

أهدي هذا الكتاب...

المؤلف



الفهرس

الفصل الثالث

رموز ومصطلحات رياضية

الفصل الرابع

مقاييس النزعة المركزية

الفصل الخامس

مقاييس التشتت

الفصل السادس

تحليل الإرتباط

الفصل السابع

تحليل الإنحدار

الملاحق

المقدمة

الحمد لله وسلامٌ على عبادهِ الذين أصطفى وصلى الله على سيدنا محمد وعلى آله وصحبهِ وسلم. أما بعد فإن مبادئ الإحصاء يعتبر من المواضيع المهمة التي تشكل المدخل الأساسي إلى علم الإحصاء وفروعه المتقدمة والأدوات المستخدمة في الكثير من العلوم المختلفة والعديد من التطبيقات العملية. وعلى هذا الأساس تضمن هذا الكتاب الأساليب الإحصائية البسيطة والمهمة في نفس الوقت والتي تشكل المادة الأساسية والمنهاج الدراسي لطلبة الإحصاء المبتدئين قبل الدخول في الأساليب الإحصائية المتقدمة التي سوف يتناولوها في المراحل المتقدمة في قسم الإحصاء فضلاً عن إمكانية إستخدامه في الفروع المختلفة من أقسام الإدارة، الإقتصاد، المحاسبة، والعلوم المالية والمصرفية نظراً لمحاولة الكاتب الإعتماد على الناحية التطبيقية من خلال أمثلة إدارية وإقتصادية عديدة والإبتعاد عن الجانب النظري أو الرياضي البحت التي لاتهم الأقسام العلمية الأخرى (غير قسم الإحصاء) من طلبة أو باحثين والتي ربما تشكل عبئاً على الطالب ويقلل من تركيزه على الجوانب العملية والتطبيقية مع مراعاة الفترة الزمنية المحدودة المخصصة لتلك الأقسام لدراسة الإحصاء فيها. كما يمكن استخدام هذه الأساليب الإحصائية كمنهاج دراسي أو أسلوب توضيحي للباحثين في جوانب علمية أخرى عديدة منها على سبيل المثال لا الحصر علوم الطب، الهندسة، الرياضيات، الإجتماع، الفلسفة، الكيمياء، الجغرافية و ...الخ.

تضمنَ الكتاب استخدام اللغة الإحصائية السهلة المخصصة لمحاكاة الطلبة المبتدئين في دراسة علم الإحصاء والإبتعاد عن المصطلحات العلمية المتقدمة التي تحتاج إلى مواضيع متقدمة لشرحها، كما حاول استخدام اللغة العربية البسيطة المفهومة متجنباً المفردات والجمل المعقدة والتي ربما تكون صعبة وغير مفهومة من قبل الدارسين والباحثين من الطلبة اللذين لايجيدون التحدث باللغة العربية بشكل كبير. وإضافة ملحق خاص بأهم المصطلحات العلمية المستخدمة في متن الكتاب وما يقابلها باللغة الإنكليزية

وملخص يتضمن أهم المعايير الإحصائية المستخدمة ومايقابلها من الصيغ الرياضية وطرائق حسابها، فـضلاً عن ترجمة هذه النسخة العربية إلى اللغة الكردية لـكي تكـون أول كتـاب مقـدم باللغـة الكرديـة في علـم الإحصاء مهداة إلى طلبتنا الإعزاء والباحثين في إقليم كوردستان العراق.

أخيراً إن الكتاب هذا بأمثلته التوضيحية الكثيرة وتركيزه على الجانب التطبيقي وتصميمه بحيث لايعتمد على معرفة متعمقة برياضيات بحته نأمل بعون الـلـه عز وجل أن يكون كتاباً جيداً بهذا الموضوع ومساعداً ودليلاً في تدريس مادة الإحصاء ووضع الأسئلة الإمتحانية وللغرض الـذي وضـع مـن أجلـه. ومـع ذلك فسيظل الكتاب بحاجة إلى كل الملاحظات التي سـترد مـن قبـل طلبتنـا الأعـزاء وزملاءنـا التدريـسيين الأفاضل لتجاوز مواطن الضعف في الكتاب لأن الكمال لله وحده عز وجل.

(فَأَمَّا الزَّبَدُ فَيَذْهَبُ جُفَاء وَأَمَّا مَا يَنفَعُ النَّاسَ فَيَمْكُثُ فِي الأَرْضِ كَذَلِكَ)

صدق الـلـه العظيم

الفصل الأول

مفهوم الإحصاء

Concept of Statistic

الفصل الأول

مفهوم الإحصاء

Concept of Statistic

1-1: علم الإحصاء:

عُرِفَ علم الإحصاء قديماً إذ وردت إشارة عن العد من قبل المؤرخ اليوناني هيرودوتس حين ذكر أنه في عام 480 ق.م. إستعمل أحد قادة الجيوش طريقة بدائية بسيطة لمعرفة عدد جيشه.

وفي القرآن الكريم وردت إشارات كثيرة تقرن الإحصاء بعملية العد، قال تعالى: [لَقَد أحْصَاهُمْ وعدّهُم عَدَا] (1)، وفي السيرة النبوية الشريفة مثال رائع لعملية العد ذلك عندما قام النبي(صلى الله عليه وسلم) بتقدير بالغ الدقة لعدد جيش قريش يوم بدر حينما علم أن قريش تنحر لجيشها كل يوم تسعٌ من الإبل.

كلمة الإحصاء (Statistics) مشتقة من الكلمة اللاتينية (Status) أو من الكلمة الإيطالية (Statista) أو من الكلمة الألمانية (Statistik) وجميعها تعني فيما تعنيه حقائق ومعلومات عن الدولة (Political State) حيث استخدم هذا المفهوم لجمع المعلومات الخاصة بافراد المجتمع لأغراض تكوين فكرة عن قوة العمل حينذاك وتكوين قاعدة معلومات من خلالها يمكن للدول فرض ضرائب لتعزيز وضعها المالي.

إن معنى الإحصاء يكون ضمنياً، ولكن كي يكون أدل، نقول إن الإحصاء عبارة عن دراسة خصائص المجتمع وتكوين إستنتاجات حول مجموعة كبيرة من البيانات عندما يكون قد تم الحصول على جزء صغير فقط من تلك البيانات.

يمكن تعريف علم الإحصاء بأنه وسيلة أو أداة يمكن من خلالها تجميع الحقائق والمعلومات وصياغتها بشكل عددي، عدّها أو تقديرها طبقاً لمستوى معقول من الدقة، وجمع هذه الحقائق والمعلومات بشكل دوري منتظم ولأغراض محددة سلفاً، ومرتبة

(1) القرآن الكريم ، سورة مريم ، آية (94).

بالشكل الذي يبين العلاقة مابينها. وبإختصار يمكن تعريف علم الإحصاء على أنه علم التقديرات والإحتمالات.

كما عرف الإحصاء بأنه العلم الذي يهتم بتوفير الحقائق الرقمية للظواهر المختلفة ومن ثم ترتيبها وعرضها ثم تحليلها للوصول إلى نتائج محددة بدقة بهدف فهم الظاهرة من جهة ووضع المقترحات المختلفة لمتابعة سيرها المستقبلي من جهة أخرى[1].

استخدم الإحصاء في مجالات عديدة مثل العلوم البايلوجية والطبية، الزراعة، الإقتصاد، الإدارة، التأمين، علم الإجتماع، علم النفس، الأجناس، الصناعة، الكيمياء، الرياضة، وغيرها.

2.1: مراحل العملية الإحصائية:

تتضمن مراحل العملية الإحصائية ما يلي:

1- **جمع البيانات:** وهي المعلومات الأولية العددية ويتم الحصول عليها من المصادر الحكومية أو الخاصة المسؤولة أو بإجراء إستفتاء أو إختيار عينة دون الحاجة إلى دراسة الكل.

2- **تنظيم البيانات:** إن البيانات التي يتم الحصول عليها تنظم عادةً بجداول إحصائية أو برسوم بيانية لغرض معالجتها رياضياً ولسهولة الإطلاع عليها ومعرفة بعض المؤشرات الأولية.

3- **المعالجة الرياضية:** إذ تتم من خلالها معالجة البيانات رياضياً وذلك لإستخراج نتائج عددية لها مؤشرات إحصائية مثل مقاييس النزعة المركزية أو التشتت أو معاملات الإرتباط وغيرها.

4- **تحليل النتائج:** وتعد من أهم مراحل العملية الإحصائية وبدونها تبقى النتائج مجرد أرقام صماء لا معنى لها ويتطلب التفسير قدراً كافياً من الأمانة والصدق

(1) د. زين العابدين عبد الرحيم البشير ، الاستدلال الإحصائي ، جامعة الملك سعود ، 1997.

وعدم التحيز والخبرة والإلمام التام بالموضوع المبحوث.

هنالك ثلاث إتجاهات من علم الإحصاء تكمل إحداها الأخرى وهي كما يلي:

1- **الطريقة الإحصائية Statistical Method:** تبين هذه الطريقة كيفية إختيار العينة Sample ثم جمع البيانات ودراسة البيانات رياضياً وإستنتاج كل ما أمكن عن البيانات، فهي مجموعة من الأساليب والمعادلات الرياضية والقوانين والإجراءات التي تساعدنا في البحث في أي موضوع إحصائي وتطبيقه في مجالات عديدة مختلفة.

2- **النظرية الإحصائية Statistical Theory:** وهي الركيزة العلمية التي تستند عليها الطريقة الإحصائية أو النظريات التي تفسر المعادلات والقوانين والأساليب التي تستعمل في الإحصاء وكيفية إشتقاقها والحصول عليها بالصيغة النهائية المعروفة إحصائياً.

3- **الإحصاء التطبيقي Applied Statistics**: وهو تطبيق أدوات الإحصاء في مجال معين ويتطلب ذلك معرفة تامة في الطريقة الإحصائية وإلماماً كبيراً في ميدان البحث الذي يبحثه مثل ميدان الإقتصاد، الإدارة، الصناعة، التربية، الإجتماع،...الخ.

3.1: أصناف علم الإحصاء:

يمكن تصنيف أقسام الإحصاء إلى قسمين رئيسين:

1- **الإحصاء الوصفي :**Descriptive Statistics

كل شيء يهتم بعملية جمع وتحليل وتفسير وتمثيل البيانات (Data) يرجع إلى علم الإحصاء كحساب معدل الإنفاق الشهري للعائلة أو جمع وعرض البيانات عن حوادث إجتماعية كالزواج والطلاق أو معرفة أراء المجتمع حول مشكلة ما من خلال إستمارة الإستبانة وحتى دراسة القوانين الضابطة لسلوك النيترونات والأكترونات. وعلى هذا الأساس يستخدم الإحصاء الوصفي لوصف الحقائق وتحويلها إلى أرقام وعرضها بشكل مناسب بإستخدام العرض البياني (Graphical Presentation) للتعبير عن البيانات الإحصائية من خلال جداول أو خرائط أو رسوم بيانية تهدف إلى إعطاء صورة عامة عن

إتجاه الظاهرة ومن هذه الرسوم (الأعمدة المستطيلة، المنحنيات والدوائر،..الخ) وكما يقال يمكن أن يغني الرسم عن ألف كلمة في توضيح الظاهرة المدروسة. من جانب آخر يتضمن الإحصاء الوصفي الدراسة الرياضية والتي يتم من خلالها حساب بعض المؤشرات الإحصائية كمقاييس النزعة المركزية مثل الوسط الحسابي، الوسيط، المنوال،..الخ ومقاييس التشتت مثل المدى، الإنحراف المعياري، التباين،..الخ فضلاً عن مقاييس العلاقة بين ظاهرتين أو أكثر مثل معامل الإرتباط والإنحدار و...الخ.

2- **الإحصاء الإستدلالي** Inference Statistical**:**

يهتم الإحصاء الإستدلالي أو التحليلي بمعالم المجتمع قيد الدراسة ويستخدم في ذلك:

أ- **التقدير** Estimation**:** يمكن تقدير معالم المجتمع قيد الدراسة من خلال التقدير النقطي (Point Estimation) مثل تقدير معدل المجتمع من خلال تقدير الوسط الحسابي أي بقيمة واحدة مقدرة من العينة لذلك المجتمع قيد الدراسة فضلاً عن التقدير بفترة (Interval Estimation) وذلك في صورة فترة محصورة بين حدين أعلى وأدنى وبدرجة ثقة أو إحتمال معين من أن معلمة المجتمع تقع ضمن هذه الفترة.

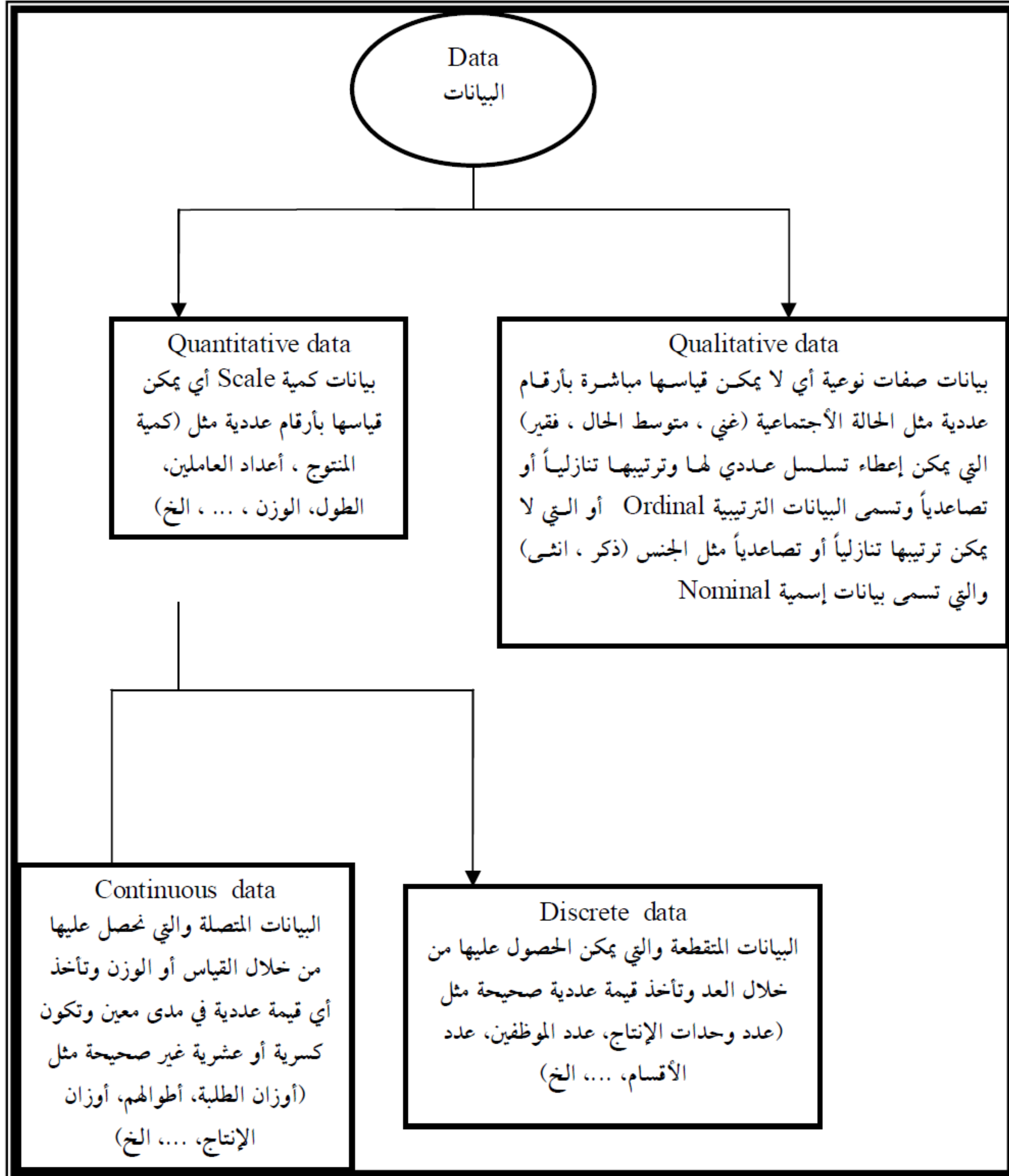
ب- **إختبار الفرضيات:**Test of Hypothesis ويعني إستخدام البيانات التي جمعت من المجتمع كعينة والمؤشرات الإحصائية المقدرة حول الظاهرة قيد الدراسة للوصول إلى قرار بشأن الفروض التي وضعت في بداية الدراسة كتفسير مؤقت لتلك الظاهرة والقرار يمكن إن يكون بقبول أو رفض الفرضية.

4.1: البيانات الإحصائية:

وهي مجموعة الحقائق والمعلومات التي تتعلق بالظاهرة قيد الدراسة وتشكل المادة الخام لعلم الإحصاء ولا يمكن تصور الطريقة الإحصائية بدونها وتقسم البيانات إلى عدة أقسام يمكن تلخيصها من خلال المخطط الآتي:

المخطط (1.1)

أقسام البيانات



وعلى هذا الأساس يمكن تعريف المتغير (Variable) بأنه عبارة عن خاصية ما تأخذ قيماً مختلفة بإختلاف الأشخاص، أو الأماكن أو الأشياء، ومتى ما حددنا طول أو وزن أو عمر شخص ما، فإن النتيجة تعتبر دائما قيمةً لذلك المتغير، وعندما ندخل عوامل الصدفة في تحديد القيم المستخرجة فإن المتغير في هذه الحالة يسمى المتغير العشوائي (Random Variable). وأن القيم الناتجة عن عمليات القياس غالباً ما يعبر عنها بإنها مشاهدات أو قياسات.

مصادر البيانات تقسم الى قسمين:

مصادر ميدانية مباشرة: حيث يتولى الباحث بنفسه توفير الحقائق والمعلومات حول الظاهرة المدروسة مثل (الإستبانة، وغيرها).

المصادر الرسمية والتاريخية: حيث تتولى المؤسسات المهتمة بتوفير البيانات الإحصائية عن مختلف الظواهر الإقتصادية والإجتماعية والصحية والثقافية والعلمية... الخ.

5.1: المجتمع الإحصائي Statistical Population:

المجتمع هو عبارة عن جميع القيم أو المفردات التي يمكن أن يأخذها المتغير والتي نرغب بالحصول على إستنتاجات حولها فهي تهم الباحث أو متخذ القرار، فمثلاً إذا كانت دراستنا متعلقة بالدخول الشهرية لموظفي كلية الإدارة والإقتصاد في جامعة صلاح الدين فإن المجتمع في هذه الحالة هو الموظفين في تلك الكلية وتلك الجامعة فقط. وإما أن يكون المجتمع محدود (Finite Population) وهو المجتمع الذي يمكن حصر عدد مفرداته كما هو الحال في الدخول الشهرية لموظفي كلية الإدارة والإقتصاد في جامعة صلاح الدين، أو أن يكون المجتمع غير محدود (Infinite Population) وهو المجتمع الذي من الصعب أو المستحيل حصر عدد مفرداته مثل الأسماك في البحر، أو النجوم في الفضاء الخارجي.

6.1: العينة Sample:

العينة هي عبارة عن جزء من المجتمع وتمثل مجموعة من المشاهدات أختيرت بطريقة ما منـه، فعندما يتعذر دراسة المجتمع ككل لأسباب تتعلق بإختصار الوقت والجهد بالإضافة إلى تخفيض الكلفـة أو إستحالة إجراء الدراسة على كافة عناصر المجتمع في بعض الحالات حيـث أن إجـراء الفحـص ينطـوي عـلى إتلاف عناصر المجتمع مثل فحص الدم أو العلب التي تتلف عند فحصها لذلك تتم الإستعاضة عـن دراسـة المجتمع بدراسة العينة وصفاتها ومنها نستطيع أن نستنتج خواص المجتمع الأصلي، الذي أخذت منه هـذه العينة.

على الرغم من المزايا التي تتحقق من خلال اللجـوء إلى اسـتخدام العينـات إلا أن إختيـار العينـات يصاحبه نوعين رئيسين من الأخطاء وهما:

1- **أخطاء الصدفة:** أو ما تسمى بأخطاء عشوائية وهي عبارة عن أخطـاء إحـصائية غـير مقـصودة وناتجة عن كون العينة جزءاً من كل المجتمع، فلا يشترط دائماً أن يمثل الجزء الكل تمثيلاً كـاملاً، ويمكن التغلب على هذه الأخطاء إما عن طريق إختيار أفضل الطرائق في عملية إختيـار العينـة أو عن طريق زيادة حجم العينة، فكلما زاد حجم العينة كلما أصبحت أكثر تمثيلاً للمجتمع.

2- **أخطاء التحيز:** وهي عبارة عن أخطاء متعمدة لأنها تحدث نتيجة تعمد من الباحث الـذي قـد يقوم بإختيار أفراد العينة بشكل متحيز نحو فئة معينة أو طبقة محددة. وهنا يمكن أن يكـون زيادة حجم العينة بشكل متحيز يؤدي إلى زيـادة حجـم الخطـأ. ويمكـن التغلـب عـلى أخطـاء التحيز عن طريق التخطيط الجيد لإختيـار وتنفيـذ العينـة. وكـذلك المراجعـة المـستمرة لمعيـار الإختبار بالإضافة إلى الإستعانة بالخبرات في هذا المجال.

تحديد حجم العينة:

عند تحديد حجم العينة يؤخذ بنظر الإعتبار مدى تجانس وحدات المجتمع ومدى الثقة التي يـود الباحث أن يلتزمها في الدراسة، فإذا كانت درجة التجانس كبيرة بين وحدات المجتمع أمكن الإكتفـاء بعينـة صغيرة الحجم أما إذا كان المجتمع غير متجانس فمن الضروري أن يكون حجم العينة كبـيراً. كـذلك يؤخـذ بنظر الإعتبار عند تحديد حجم العينة الإمكانيات المادية والوقت المحدد للدراسة الذي قد يسبب بإختيـار حجم عينة أقل من الحجم المناسب.

طرائق المعاينة:

إن اسلوب إختيار العينة يسمى بالمعاينة (Sampling) إذ تمثل العينة مجموعة وحـدات المعاينـة (Sampling Units) التي يتم إختيارها من إطـار عـام يـسمى إطـار المعاينـة (Sampling Frame) وقـد تـضم أفراداً أو أُسَراًّ أو وحدات من المنتوج أو عبوات تضم أكثر من وحدة أي أن وحدة المعاينة قد تتكون مـن عنصر أو أكثر، وتعطي العينات نتائج عالية الدقـة تتناسـب طردياً مـع درجـة تمثيـل البيانـات للمجتمـع. وهنالك مجموعتين رئيسيتين من طرائق إختيار العينات وهما:

أولاً: طرائق المعاينة العشوائية:

وهي الطرائق المتعلقة بإختيار أفـراد العينـة عـشوائياً بمعنـى أن لكـل مفـردة في المجتمـع فرصـة مساوية لفرصة أي مفردة أخرى في الظهور بالعينة. فمثلاً إذا كان عدد أفـراد المجتمـع (1000) فـإن فرصـة ظهور أي فرد في العينة تساوي واحد من (1000)، ومن أهم أنواع العينات العشوائية:

1- **العينة العشوائية البسيطة** Simple Random Sample:

تُعَدْ من أبسط أنواع العينات وذلك عندما تكون وحدات مجتمع الدراسـة في حالة تماثل إذ تتاح لكل وحدة من وحدات المجتمع فرصة الظهور في العينة بـإحتمال متساوي ومستقل ومن بين الطرائق التي تستخدم في الإختيـار العـشوائي هـو طريقـة (القرعة)، وذلك بكتابة أرقام أفراد المجتمع الأصلي على بطاقات صغيرة وبعـد خلطهـا

بشكل جيد يتم سحب عدد منها بطريقة عشوائية غير متحيزة تمثل عينة البحث.

أو استخدام جداول الأعداد العشوائية وبخاصةً عندما يكون حجم العينة كبيراً وهي تتكون من أعمدة وصفوف حيث يتم إختيار صف وعمود بشكل عشوائي من الجدول ثم يتم إختيار عدد الخانات الرقمية المناسبة لأكبر رقم في المجتمع ويتم إختيار الأفراد حسب تسلسل الأرقام الموجودة في الجدول بشكل أفقي وعمودي إلى أن يكتمل العدد المطلوب من عينة البحث.

هنالك عدة طرائق لتحديد حجم العينة العشوائية البسيطة منها:

أ- تحديد حجم العينة على أساس المتوسط: يتم ذلك من خلال تطبيق الصيغة الآتية:

$$n = \frac{z^2 \sigma^2}{e^2} \qquad \dots \quad (1.1)$$

n: تمثل حجم العينة.

z: القيمة المعيارية لمتغير معين فإذا كانت تحت مستوى ثقة (95%) فإنها تساوي (1.96).

σ^2 : تباين المجتمع.

e: الخطأ المسموح به أي أقل إنحراف مسموح به للمتوسط الحسابي للعينة عن المتوسط الحسابي للمجتمع.

مثال ذلك: إذا كان تباين مجتمع البحث يساوي (5) ومستوى الثقة المطلوب (95%) وأن الخطأ المسموح به يساوي (0.01)، فإن حجم العينة المطلوب:

$$n = \frac{(1.96)^2 (5)}{(0.01)^2} = 192080$$

ب- تحديد حجم العينة على أساس النسب: يتم ذلك من خلال تطبيق الصيغة

الآتية:

$$n = \frac{z^2 \, P(1-P)}{e^2} \qquad \ldots \quad (1.2)$$

حيث تمثل P نسبة النجاح المفترضة في المجتمع.

مثال ذلك: لو إفترضنا أن من بين كل عشرة إتصالات مع العملاء تنجح عملية واحدة في توقيع عقد البيع. ما هو حجم العينة إذا كان مستوى الثقة المطلوب (95%) والخطأ المسموح به (0.05).

$$n = \frac{(1.96)^2 (0.10)(1-0.10)}{(0.05)^2} \approx 138$$

2- **العينة العشوائية الطبقية** Stratified Random Sample:

يتم تقسيم المجتمع الأصلي تحت الدراسة إلى عدة طبقات غير متجانسة في حين تحتوي كل طبقة على مفردات متجانسة فيما بينها ومن ثم يختار بطريقة عشوائية من هذه الطبقات العدد المطلوب من المفردات بما يتناسب مع حجم كل طبقة من طبقات المجتمع. وتؤدي هذه الطريقة التناسبية إلى زيادة تمثيل العينة وتمكن الباحث من استخدام عينات أصغر مما يقلل من التكاليف الخاصة بالبحث.

3- **العينة العشوائية ذات المراحل المتعددة** Multi-stage Random Sample:

وهي أن تكون المعاينة في مجموعات وتسحب عينة عشوائية بسيطة من المجموعات المكونة للمجتمع المراد دراسته ومن ثم القيام بتعداد شامل لكل مفردات المجموعات التي وقع عليها الإختيار أو تقوم بمعاينة مفردات تلك المجموعات المختارة (معاينة ذات مرحلتين) وقد تضيف مراحلاً أخرى من المعاينة حسب طبيعة البحث وتسمى حينذاك تلك المعاينة بالمعاينة المتعددة المراحل. فعند القيام ببحث لزراعة محصول ما قد تأخذ القرى كوحدات إبتدائية ثم نختار منها المزارع كوحدات ثانوية ثم نعاين قطعاً من المزارع

المختارة وهذه معاينة ذات ثلاث مراحل وهكذا.

4- **العينة العشوائية المنتظمة** Systematic Random Sample:

تستخدم هذه الطريقة في حالة تجانس المجتمع الأصلي أو عشوائيته ويتم إختيار المفـردات مـن مسافات متساوية على القائمة إذ نختار الوحدة الأولى للعينـة بطريقـة عـشوائية وهـي التـي تحـدد بـاقي مفردات العينة من خلال إختيارها على شكل مسافات متساوية، فمثلاً أردنا إختيار عينة مـن طلبـة قـسم الإقتصاد بنسبة (10%) من الطلبة وكان عـددهم (500) طالـب، ففـي هـذه الحالـة يكـون حجـم العينـة المطلوب (500 ×0.10) ويساوي (50) طالب ويعتمد إختيار كل فرد من الأفراد على حجم العينة وحجـم المجتمع فقد يكون (1-8) أو (1-15)...الخ. وإذا ما أردنا أن نأخذ طالباً واحداً مـن كـل (10) طـلاب فإننـا نختار الأول عشوائياً ثم عند إختيارنا الثاني نضيف (10) إلى الرقم الأول وهكذا فإذا كان رقم الطالب الأول (6) فيكون الطالب صاحب هذا الرقم هو الأول في العينة ثم نضيف العدد (10) إلى الرقم الذي أخترناه ثم إلى الذي يليه وهكذا حتى نصل إلى الحجـم المطلـوب للعينـة فتكـون الأرقـام التـي تـم إختيارهـا (6، 16، 26،...الخ).

ثانياً: طرائق المعاينة غير العشوائية:

يتم من خلالها إختيار وحدات المعاينة إما بالصدفة أو أن تكون ذات إختيار متعمد لغـرض إجـراء دراسة محددة، لذا لاتخضع عملية الإختيار للقرعة أو لمجرد الحظ، ومن هذه الأنواع:

1- العينة بالمصادفة (الصدفة) Accidental Sample:

وهنا لا يكون للباحث أي تدخل في إختيار العينة فإذا ما أراد الباحـث التعـرف عـلى الـرأي العـام تجاه قضية ما فإنه ينزل إلى الشارع ويسأل أول شخص يصادفه من الأشخاص وهكذا...

2- العينة الحصصية Quota Sample:

يتم من خلالها تقسيم المجتمع إلى فئات ويختار من كل فئة مجموعة من الأفراد ممثلةً له وهنا يتم الإختيار حسب مايراه مناسباً وليس بشكل عشوائي، فعندما يقسم الباحث المجتمع إلى فئات ذات خصائص معينة مثل (طلاب، مدرسون، عمال،...الخ)، ثم يختار من كل فئة مجموعة تمثلة دون أن يكون الإختيار عشوائياً، وهنا يكون مجال الإختلاف الوحيد بين العينة الطبقية والعينة الحصصية في مجال إختيار الأفراد فقط.

3- العينة العمدية Purposive Sample:

لإستخدام هذه الطريقة يجب أن يتوفر للباحث معرفة المؤشرات الإحصائية للمجتمع الأصلي وللوحدات التي يرغب في إختيارها وفي ضوء تلك المعرفة يقوم بإختيار وحدات معينة يعتقد أنه تمثل المجتمع الأصلي فإذا أراد الباحث دراسة الرأي العام حول قضية معينة فأنه يختار من رجال السياسة والآخرين عدداً من الأفراد لإجراء دراسته، وهذا النوع من العينات لا يمثل المجتمع إنما يمثل رأي الأفراد الداخلين في العينة فقط.

تمارين الفصل الأول

1-1 : عرف الإحصاء وما هو أنواعه.

2-1 : عدد مراحل العملية الإحصائية.

3-1 : ما الفرق بين الإحصاء الوصفي والإحصاء الإستدلالي؟

4-1 : ما أهمية الإحصاء للعلوم الإدارية والإقتصادية؟

5-1 : ما الفرق بين العينة والمجتمع؟

6-1 : ماهي طريقة المعاينة الأكثر استخداماً في الحياة العملية؟ إشرحها؟

7-1 : إذا كان تباين مجتمع البحث يساوي **(25)** ومستوى الثقة المطلوب (99%) وأن الخطأ المسموح به يساوي **(0.05)**، فما هو حجم العينة المناسب ؟ بإستخدام العينة العشوائية البسيطة على أساس المتوسط.

8-1 : إذا كانت نسبة النجاح لطلبة الإقتصاد لمجتمع البحث يساوي **(0.80)** ومستوى الثقة المطلوب (95%) وأن الخطأ المسموح به يساوي **(0.10)**، قدر حجم العينة المناسب؟ بإستخدام العينة العشوائية البسيطة على أساس النسب.

9-1 : رغب أحد الباحثين بإختيار عينة نسبتها 20% من إجمالي عدد الموظفين البالغ (5000) موظف. كيف يتم ذلك وفقاً لطريقة العينة العشوائية المنتظمة؟

الفصل الثاني

تبويب وعرض البيانات

Data Tabulation and Presentation

الفصل الثاني

تبويب وعرض البيانات

Data Tabulation and Presentation

1.2: مقدمة Introduction:

البيانات التي يتم جمعها بإسلوب العينة أو غيره بدون فائدة إن لم يتم تنظيمها وجدولتها ، لـذا تعتبر عملية تبويب البيانات من الخطوات الأساسية لعملية تحليل البيانات والحصول على نتائج ذات قيمة من خلال تصنيف البيانات وترتيبها وعرضها وبما يؤدي إلى إعطاء صورة واضحة لها وإلى فهم خصائصها ومكوناتها ويمهد الطريق للحصول على نتائج أكثر عمقاً وتأثيراً في المراحل التحليلية اللاحقة. وعلى هذا الأساس سوف يتم التركيز في هذا الفصل على أساليب تبويب البيانات في جداول خاصة تدعى بجداول التوزيعات التكرارية، كذلك إستعراض عرض البيانات هندسياً.

2.2: تبويب وعرض البيانات Data Tabulation and Presentation:

سوف يتم من خلال هذه الفقرة تحويل البيانات الخام أو الأولية التي تم جمعها باستخدام العينة والتي تسمى عادة بيانات غير مبوبة إلى بيانات مبوبة بالإعتماد على جداول خاصة تدعى بالتوزيعات التكرارية التي يتم من خلالها عرض البيانات المصنفة.

1.2.2: التوزيع التكراري Frequency Distribution:

التوزيع التكراري يمثل تلخيص وترتيب البيانات التي جمعت وصنفت، وقسمت إلى عدد من المجاميع كل منها تسمى الفئة (Class). هذه الفئات تكون مرتبة تصاعدياً أو تنازلياً حسب طبيعة البيانات. ويسمى توزيع عدد قيم x حسب الفئات (التوزيع التكراري) وقد تكون فئات التوزيع التكراري متساوية في الطول أو غير متساوية.

أفرض لدينا $\left(x_1, x_2, \ldots, x_n\right)$ تمثل عينة عشوائية حجمها n من المشاهدات ونرغب تلخيص هذه البيانات في توزيع تكراري عدد فئاته هو m، ولدينا ما يلي:

- **المدى الكلي** Total Range: المدى الكلي هو عبارة عن الفرق مابين أكبر قيمة x_L وأصغر قيمة x_S في المجموعة مضافاً له العدد واحد، أي أن:

$$T.R = x_L - x_S + 1 \qquad \cdots \qquad (2.1)$$

- **عدد الفئات** Number of Classes: عدد الفئات هي عبارة عن عدد المجاميع التي يتألف منها الجدول التكراري، وهنالك عدة صيغ يمكن من خلالها تحديد عدد فئات الجدول، منها طريقة يول Yule وهي:

$$m = (2.5)\cdot\sqrt[4]{n} = (2.5)n^{1/4} \qquad \cdots \qquad (2.2)$$

فإذا كانت n = 50 مثلاً فإن عدد الفئات هي:

$$m = (2.5)\cdot\sqrt[4]{50} = (2.5)(50)^{1/4}$$
$$= (2.5)(2.659) = 6.6475 \cong 7$$

أو إستخدام صيغة سترجس Sturges وهي:

$$m = 1 + (3.322)\log_{10}(n) \qquad \cdots \qquad (2.3)$$

فإذا كانت n = 50 مثلاً فإن عدد الفئات حسب الصيغة أعلاه تكون كما يلي:

$$m = 1 + (3.322)\log_{10}(50)$$
$$= 1 + (3.322)(1.69897) = 6.644 \cong 7$$

حيث يتم تقريب الناتج إلى أقرب عدد صحيح عند التطبيق العملي لإن من المنطقي أن يكون عدد الفئات عدد صحيح.

- **طول الفئة** Length of Class: وهوعبارة عن مقدار المسافة مابين الحد الأدنى للفئة وحدها الأعلى. وأن طول الفئة L يتناسب عكسياً مع عدد الفئات فكلما زاد طول الفئة قل عدد الفئات والعكس صحيح. وفي حالة الجداول ذات الفئات المتساوية

بالطول يمكن تحديد طول الفئة L من خلال الصيغة الآتية:

$$L = \frac{T.R}{m} \qquad \cdots \qquad (2.4)$$

- **الحد الأدنى والأعلى للفئة** Lower and Upper Limit of a Class**:** لكل جدول تكراري بداية ونهاية. فالبداية تعني الحد الأدنى للفئة L.L في حين تمثل النهاية الحد الأعلى U.L لها. ويمكن تكوين حدود الفئات في حالة تساوي أطوالها بالشكل الآتي:

تسلسل الفئة	الحد الأدنى	الحد الأعلى
1	x_S	$x_S + L$
2	$x_S + L$	$x_S + 2L$
3	$x_S + 2L$	$x_S + 3L$
$\vdots$	$\vdots$	$\vdots$
m	$x_S + (m-1)L$	$x_S + mL$

ملاحظة: الجدول السابق مناسب للبيانات المستمرة أما إذا كانت البيانات من النوع المتقطع فنطرح العدد واحد من الحد الأعلى للفئة.

- **مركز الفئة** Center of a Class**:** قيم x التي تتوسط المسافة بين الحد الأدنى والأعلى للفئة تمثل مراكز الفئات، أي أن:

$$x = \frac{L.L + U.L}{2} \qquad \cdots \qquad (2.5)$$

- **تكرار الفئة** Class Frequency**:** وهو عبارة عن عدد المشاهدات التي تقع ضمن تلك الفئة بحيث أن مجموعها يمثل حجم العينة n فإذا رمزنا لتكرارات الفئات بالرموز

فإن: $\left(f_1, f_2, \ldots, f_m\right)$

$$\sum_{i=1}^{m} f_i = f_1 + f_2 + \cdots + f_m = n$$

من الملاحظ أنه ليس من الضروري أن يكون الحد الأدنى للفئة الأولى مساوٍ تماماً لأصغر قيمة في المجموعة بل قد تكون أقل منها، كذلك الحد الأعلى للفئة الأخيرة ليس بالضروري أن يساوٍ أكبر قيمة بل قد تكون أكبر منه وذلك لتسهيل العمليات الحسابية. كذلك فإن الجدول التكراري قد يكون مغلق (يمتلك حداً أدنى للفئة الأولى وحداً أعلى للفئة الأخيرة) أو مفتوح (لا يمتلك حداً أدنى للفئة الأولى وحداً أعلى للفئة الأخيرة أو كليهما معاً) وذلك يعتمد هو الآخر على طبيعة الدراسة.

مثال (1.2)

تم دراسة أجور العاملين اليومية للمواد الإنشائية وكانت كمايلي:

29	38	37	35	30	26	41	37	34
30	44	42	37	33	31	27	40	38
32	28	49	40	39	34	30	39	35
39	31	33	26	44	31	31	46	43
38	35	35	32	45	36	32	34	48

المطلوب: نظم هذه البيانات في جدول تكراري

الحل:

$n = 45$ فإن عدد الفئات m هي كما يلي:

$$m = (2.5) \cdot \sqrt[4]{45} = (2.5)(45)^{1/4}$$
$$= (2.5)(2.59) = 6.475 \cong 6$$

من الجدول السابق نلاحظ أن أصغر قيمة $x_S = 26$ وأكبر قيمة $x_L = 49$ لذلك فإن المدى الكلي هو:

$$T.R = x_L - x_S + 1 = 49 - 26 + 1 = 24$$

لذلك فإن طول الفئة:

$$L = \frac{T.R}{m} = \frac{24}{6} = 4$$

وعلى هذا الأساس يمكن تكوين الجدول الآتي:

تسلسل الفئة	الحد الأدنى للفئة	الحد الأعلى للفئة	حدود الفئات بالشكل النهائي
1	26	29	29-26
2	30	33	33-30
3	34	37	37-34
4	38	41	41-38
5	42	45	45-42
6	46	49	49-46

بعد تحديد الفئات نقوم بتفريغ البيانات حسب الفئات وكما يلي:

تسلسل الفئة	الفئات (الأجور)	التكرارات (بالإشارات عدد العمال)	التكرار f_i
1	29-26	$\cancel{||||}$	5
2	33-30	$\cancel{||||}$ $\cancel{||||}$ //	12
3	37-34	$\cancel{||||}$ $\cancel{||||}$ /	11
4	41-38	$\cancel{||||}$ ////	9
5	45-42	$\cancel{||||}$	5
6	49-46	///	3

يمكن تلخيص النتائج مع مراكز الفئات من خلال الجدول الآتي:

تسلسل الفئة	الفئات (الأجور)	التكرار f_i (عدد العمال)	مراكز الفئات
1	29-26	5	27.5
2	33-30	12	31.5
3	37-34	11	35.5
4	41-38	9	39.5
5	45-42	5	43.5
6	49-46	3	47.5

مثال (2.2)

البيانات التالية تمثل كميات إنتاج الحنطة في الدونم الواحد إلى (60) قطعة زراعية، والمطلوب تكوين جدول تكراري فضلاً عن حساب مراكز الفئات.

23.39	32.48	35.37	34.65	20.30	26.96	42.81	17.83	24.34	34.56
35.02	22.44	43.82	37.89	25.33	35.81	72.77	40.90	67.85	65.98
61.22	28.65	42.69	40.43	43.59	63.84	34.50	23.49	34.65	23.45
58.89	50.11	35.43	23.56	55.44	56.41	39.61	24.36	32.54	43.67
72.38	61.35	73.05	32.54	23.45	62.36	39.52	34.34	32.87	54.28
78.89	52.14	35.43	43.51	50.42	36.51	31.31	28.16	22.56	37.67

الحل:

$n = 60$ فإن عدد الفئات m هي كما يلي:

$$m = (2.5)\cdot\sqrt[4]{60} = (2.5)(60)^{1/4}$$
$$= (2.5)(2.7832) = 6.9579 \cong 7$$

من الجدول السابق نلاحظ أن أصغر قيمة $x_S = 17.83$ وأكبر قيمة $x_L = 78.89$ لذلك

فإن المدى الكلي هو:

$$T.R = x_L - x_S + 1 = 78.89 - 17.83 + 1 = 62.06$$

لذلك فإن طول الفئة:

$$L = \frac{T.R}{m} = \frac{62.06}{7} = 8.8657$$

وعلى هذا الأساس يمكن تكوين الجدول الآتي:

تسلسل الفئة	الحد الأدنى للفئة	الحد الأعلى للفئة	حدود الفئات بالشكل النهائي	يمكن كتابتها للسهولة بالشكل الآتي
1	17.83	26.70	26.70-17.83	- 17.83
2	26.70	35.56	35.56-26.70	-26.70
3	35.56	44.42	44.42-35.56	-35.56
4	44.42	53.28	53.28-44.42	-44.42
5	53.28	62.15	62.15-53.28	-53.28
6	62.15	71.02	71.02-62.15	-62.15
7	71.02	79.88	79.88-71.02	79.88-71.02

بعد تحديد الفئات نقوم بتفريغ البيانات حسب الفئات وكما يلي:

تسلسل الفئة	الفئات (كمية إنتاج الحنطة بالطن)	التكرارات (بالإشارات عدد قطع الأراضي)	التكرار f_i
1	26.70-17.83	/	1
2	35.56-26.70	//// ///// /////	14
3	44.42-35.56	///// ///// ///// /////	20
4	53.28-44.42	/// /////	8
5	62.15-53.28	// /////	7
6	71.02-62.15	/ /////	6
7	79.88-71.02	////	4

يمكن تلخيص النتائج مع مراكز الفئات من خلال الجدول الآتي:

تسلسل الفئة	الفئات (كمية إنتاج الحنطة بالطن)	التكرار f_i (عدد قطع الأراضي)	مراكز الفئات
1	26.70-17.83	1	22.27
2	35.56-26.70	14	31.13
3	44.42-35.56	20	39.00
4	53.28-44.42	8	48.86
5	62.15-53.28	7	57.73
6	71.02-62.15	6	66.59
7	79.88-71.02	4	75.46

من خلال الجدول أعلاه نلاحظ أن هنالك قطعة واحدة كمية إنتاج الحنطة بالطن يتراوح مابين 17.83 طن وأقل من 26.70 طن وأن هنالك 14 قطعة زراعية كمية إنتاجها من الحنطة بالطن يتراوح مابين 26.70 طن وأقل من 35.56 طن وهكذا البقية...، كما نلاحظ من مراكز الفئات أن هنالك قطعة واحدة منتوجها بالمتوسط 22.27 طن وأن هنالك 14 قطعة زراعية منتوجها بالمتوسط 31.13 طن وهكذا البقية... كما نلاحظ من خلال المثالين السابقيين أن مجموع التكرارات يساوي عدد المشاهدات الأصلية n.

إن إسلوب التبويب الذي تم إعتماده في المثالين أعلاه يفترض حالة تساوي أطوال الفئات ولكن هنالك حالات أخرى ربما تكون أطوال الفئات غير متساوية وذلك حسب طبيعة الدراسة ومتطلباتها ولذلك يكون الأسلوب أعلاه غير مفيد في تلك الحالة، مما يتطلب تحديد عدد الفئات وحدودها على نمط التبويب الذي يحقق هدف الدراسة مع مراعاة نوع المتغير متقطعاً كان أو مستمراً.

2.2.2: التوزيع التكراري النسبي Proportionate Frequency Distribution:

إن التكرارات النسبية هي عبارة عن التكرار الإعتيادي f_i معبر عنها بنسب مئوية يمكن الحصول عليها من قسمة تكرار كل فئة على مجموع التكرارات الكلية n. أي أن:

$$f_i^* = \frac{f_i}{\text{n}}.100 \qquad \cdots \quad (2.6)$$

مثال (3.2)**: جد التوزيع التكراري النسبي للمثال السابق** (2.2)**.**

الحل:

يمكن الحصول على التوزيع التكراري النسبي من خلال تقسيم تكرار كل فئة على مجموع التكرارات الكلية 60، أي أن:

$$f_i^* = \frac{f_i}{\text{n}}.100$$

$$f_1^* = \frac{f_1}{\text{n}}.100 = \frac{1}{60}.100 = 1.667$$

$$f_2^* = \frac{f_2}{\text{n}}.100 = \frac{14}{60}.100 = 23.333$$

$$\vdots$$

$$f_7^* = \frac{f_7}{\text{n}}.100 = \frac{4}{60}.100 = 6.667$$

ويمكن تلخيصها من خلال الجدول الآتي:

الفئات (كمية إنتاج الحنطة بالطن)	التكرار f_i (عدد قطع الأراضي)	التكرار النسبي f_i^* (عدد قطع الأراضي)
26.70-17.83	1	1.667
35.56-26.70	14	23.333

33.333	20	44.42-35.56
13.334	8	53.28-44.42
11.667	7	62.15-53.28
10	6	71.02-62.15
6.667	4	79.88-71.02
100	60	المجموع

وهذا يعني أن 1.667% من قطع الأراضي يتراوح إنتاجهم مابين 17.83 طن إلى أقل من 26.70 طن وأن 23.333% من قطع الأراضي يتراوح إنتاجهم مابين 26.70 طن إلى أقل من 35.56 طن، وهكذا البقية.....، علماً أن مجموع التكرار النسبي يجب أن يساوي 100

3.2.2: التوزيع التكراري المتجمع Cumulative Frequency Distribution:

وهو التوزيع الذي يبين كمية التكرار المتجمع عند قيمة معينة من قيم المشاهدات وهو على نوعين وهما:

أولاً: التوزيع التكراري المتجمع الصاعد: وهو يبين تراكم التكرارات إبتداءاً من الفئة الأولى للجدول وإنتهاءاً بالفئة الأخيرة منه ويتم حساب التكرارات المتجمعة على أساس الحدود العليا للفئات، أي أن:

$$
\begin{aligned}
F_1 &= f_1 \\
F_2 &= f_1 + f_2 \\
&\vdots \\
F_m &= f_1 + f_2 + \cdots + f_m \qquad \cdots \quad (2.7)
\end{aligned}
$$

ويمكن تحويل توزيع متجمع صاعد F_i إلى توزيع تكراري صاعد نسبي F_i^* وذلك من خلال الصيغة الآتية:

$$F_i^* = \frac{F_i}{n}.100 \qquad \cdots \quad (2.8)$$

مثال (4.2)

بالإعتماد على المثال (1.2) **جد التوزيع التكراري المتجمع الصاعد والنسبي؟**

الحل: من خلال المثال (1.2) **لدينا الجدول الآتي:**

تسلسل الفئة	الفئات (الأجور)	التكرار f_i (عدد العمال)
1	29-26	5
2	33-30	12
3	37-34	11
4	41-38	9
5	45-42	5
6	49-46	3

التوزيع التكراري المتجمع الصاعد هو كما يلي:

$$F_1 = f_1 = 5$$
$$F_2 = f_1 + f_2 = 5 + 12 = 17$$
$$\vdots$$
$$F_m = f_1 + f_2 + \cdots + f_m = 5 + 12 + \cdots + 3 = 45$$

التوزيع التكراري المتجمع الصاعد النسبي هو كما يلي:

$$F_i^* = \frac{F_i}{n}.100$$

$$F_1^* = \frac{F_1}{n}.100 = \frac{5}{45}.100 = 11.111$$

$$F_2^* = \frac{F_2}{n}.100 = \frac{17}{45}.100 = 37.778$$

$$\vdots$$

$$F_6^* = \frac{F_6}{n}.100 = \frac{45}{45}.100 = 100$$

ويمكن تلخيص النتائج من خلال الجدول الآتي:

تسلسل الفئة	الفئات (الأجور)	التكرار f_i (عدد العمال)	التكرار المتجمع الصاعد F_i	التكرار المتجمع الصاعد النسبي $F_i^*\%$
1	29-26	5	5	11.111
2	33-30	12	17	37.778
3	37-34	11	28	62.222
4	41-38	9	37	82.222
5	45-42	5	42	93.333
6	49-46	3	45	100

وهذا يعني أن عدد العمال الذين أجورهم 37 ألف دينار فأقل هو 28 عامل وأن عدد العمال الذين أجورهم 45 ألف دينار فأقل هو 42 عامل. وأن نسبة العمال الذين أجورهم 37 ألف دينار فأقل هي 62.222% وأن نسبة العمال الذين أجورهم 45 ألف دينار فأقل هي 93.333%.

ملاحظة: إذا كانت البيانات من النوع المتقطع (كالمثال أعلاه) فإن التكرار المتجمع

الصاعد والنسبي يكون مساوي وأقل من الحد الأعلى للفئة، في حين إذا كانت البيانات من النوع المستمر فإن التكرار المتجمع الصاعد والنسبي يكون أقل من الحد الأعلى للفئة.

ثانياً: التوزيع التكراري المتجمع النازل: وهو يبين تناقص التكرارات إبتداءاً من الفئة الأولى للجدول وإنتهاءاً بالفئة الأخيرة منه ويتم حساب التكرارات المتجمعة على أساس الحدود الدنيا للفئات، أي أن:

$$
\begin{aligned}
F'_1 &= \mathrm{n} \\
F'_2 &= \mathrm{n} - f_1 \\
&\vdots \\
F'_{\mathrm{m}} &= \mathrm{n} - f_1 - f_2 - \cdots - f_{\mathrm{m}-1} = f_{\mathrm{m}} \qquad \cdots \quad (2.9)
\end{aligned}
$$

ويمكن تحويل توزيع متجمع نازل F'_i إلى توزيع تكراري نازل نسبي F'^*_i وذلك من خلال الصيغة الآتية:

$$F'^*_i = \frac{F'_i}{\mathrm{n}}.100 \qquad \cdots \qquad (2.10)$$

مثال (5.2)

بالإعتماد على المثال (1.2) **جد التوزيع التكراري المتجمع النازل والنسبي؟**

تسلسل الفئة	الفئات (الأجور)	التكرار f_i (عدد العمال)	مراكز الفئات
1	29-26	5	27.5
2	33-30	12	31.5
3	37-34	11	35.5
4	41-38	9	39.5
5	45-42	5	43.5
6	49-46	3	47.5

الحل:

التوزيع التكراري المتجمع النازل يمكن أن نحصل عليه من خلال مايلي:

$$F'_1 = \mathrm{n} = 45$$

$$F'_2 = \mathrm{n} - f_1 = 45 - 5 = 40$$

$$\vdots$$

$$F'_{\mathrm{m}} = \mathrm{n} - f_1 - f_2 - \cdots - f_{\mathrm{m}-1} = 45 - 5 - 12 - \cdots - 5 = 3 = f_{\mathrm{m}}$$

التوزيع التكراري المتجمع النازل النسبي يمكن أن نحصل عليه من خلال مايلي:

$$F_i'^{*} = \frac{F_i'}{\mathrm{n}}.100$$

$$F_1'^* = \frac{F_1'}{n}.100 = \frac{45}{45}.100 = 100$$

$$F_2'^* = \frac{F_2'}{n}.100 = \frac{40}{45}.100 = 88.889$$

$$\vdots$$

$$F_6'^* = \frac{F_6'}{n}.100 = \frac{3}{45}.100 = 6.667$$

ويمكن تلخيص النتائج من خلال الجدول الآتي:

تسلسل الفئة	الفئات (الأجور)	التكرار f_i (عدد العمال)	التكرار المتجمع النازل F_i'	التكرار المتجمع النازل النسبي $F_i'^*\%$
1	29-26	5	45	100
2	33-30	12	40	88.889
3	37-34	11	28	62.222
4	41-38	9	17	37.778
5	45-42	5	8	17.778
6	49-46	3	3	6.667

وهذا يعني أن عدد العمال الذين أجورهم الأسبوعية 34 ألف دينار فأكثر هو 28 عامل وأن عدد العمال الذين أجورهم الأسبوعية 42 ألف دينار فأكثر هو 8 عمال، وأن نسبة العمال الذين أجورهم 34 ألف دينار فأكثر هي 62.222% وأن نسبة العمال الذين أجورهم 42 ألف دينار فأكثر هي 17.778%.

كذلك يمكن تكوين توزيع تكراري متجمع صاعد أو نازل في حالة المتغيرات النوعية وذلك من خلال تجميع أوتنقيص التكرارات حسب مستويات ذلك المتغير وكما هو موضح من خلال المثال الآتي:

مثال (6.2)

فيما يلي عينة عشوائية تتكون من 50 شخص حسب تحصيلهم العلمي:

تحصيلهم العلمي	عدد الأشخاص
يقرأ ويكتب	7
إبتدائية	11
متوسطة	15
إعدادية	9
كلية	8

المطلوب تكوين جدول التوزيع التكراري الصاعد والنازل؟

الحل:

يتم تجميع التكرارات إبتداءاً من المستوى الأول (يقرأ ويكتب) وكما يلي:

تحصيلهم العلمي	**عدد الأشخاص** f_i	**اسلوب التجميع**	F	**اسلوب التنقيص**	F'
يقرأ ويكتب	7	يقرأ ويكتب	7	كل المستويات	50
إبتدائية	11	يقرأ ويكتب وإبتدائية	18	إبتدائية، متوسطة، إعدادية وكلية	43
متوسطة	15	يقرأ ويكتب وإبتدائية ومتوسطة	33	متوسطة، إعدادية وكلية	32
إعدادية	9	يقرأ ويكتب، إبتدائية ، متوسطة وإعدادية	42	إعدادية وكلية	17
كلية	8	كل المستويات	50	كلية	8

3.2: العرض البياني للبيانات:

تعتبر وسائل العرض البياني (أشكال ورسوم بيانية) ضمن أدوات الإحصاء الوصفي وذلك لتنظيم وتلخيص وعرض البيانات المبوبة أو غير المبوبة إما بديلاً عن الجداول التكرارية أو إستكمالاً لها وتمتاز هذه الوسائل بالبساطة والفعالية في عرض البيانات وإعطاء صورة واضحة (فكرة سريعة) عنها، ومن هذه الوسائل المستخدمة:

أولاً: في حالة البيانات المبوبة لدينا ما يلي:

1- المدرج التكراري Histogram:

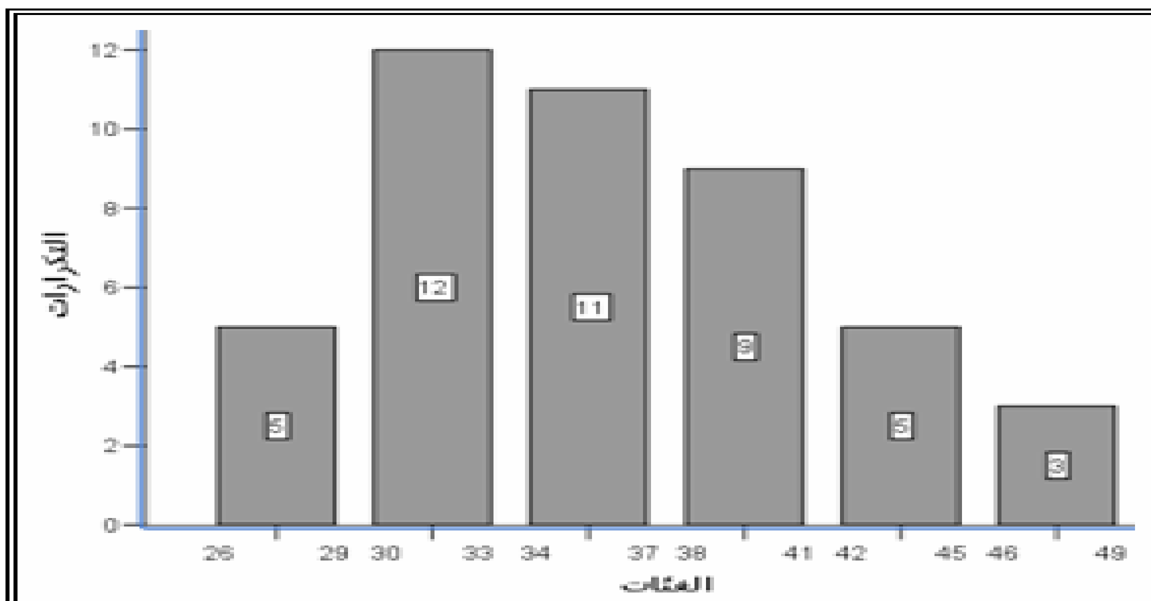
وهو تمثيل البيانات بإستخدام مجموعة من المستطيلات المتلاصقة في حالة البيانات المستمرة وغير المتلاصقة في حالة البيانات المتقطعة بحيث تكون قاعدة كل منها تساوي طول الفئة وإرتفاعها يساوي تكرارها (أو التكرار النسبي) وهو أسهل طريقة لتمثيل التوزيعات التكرارية ومساحته تتناسب مع التكرار الكلي للتوزيع إذ أن مجموع مساحات المستطيلات يكون مساوياً للمجموع الكلي للتكرارات.

مثال (7.2)

استخدم المدرج التكراري في تمثيل البيانات المبوبة في المثال (1.2):

تسلسل الفئة	الفئات (الأجور)	التكرار f_i (عدد العمال)
1	29-26	5
2	33-30	12
3	37-34	11
4	41-38	9
5	45-42	5
6	49-46	3

الحل: بما أن البيانات من النوع المتقطع لذلك فإن المدرج التكراري يتكون من مستطيلات غير متلاصقة، كما يوضحه الشكل الآتي:



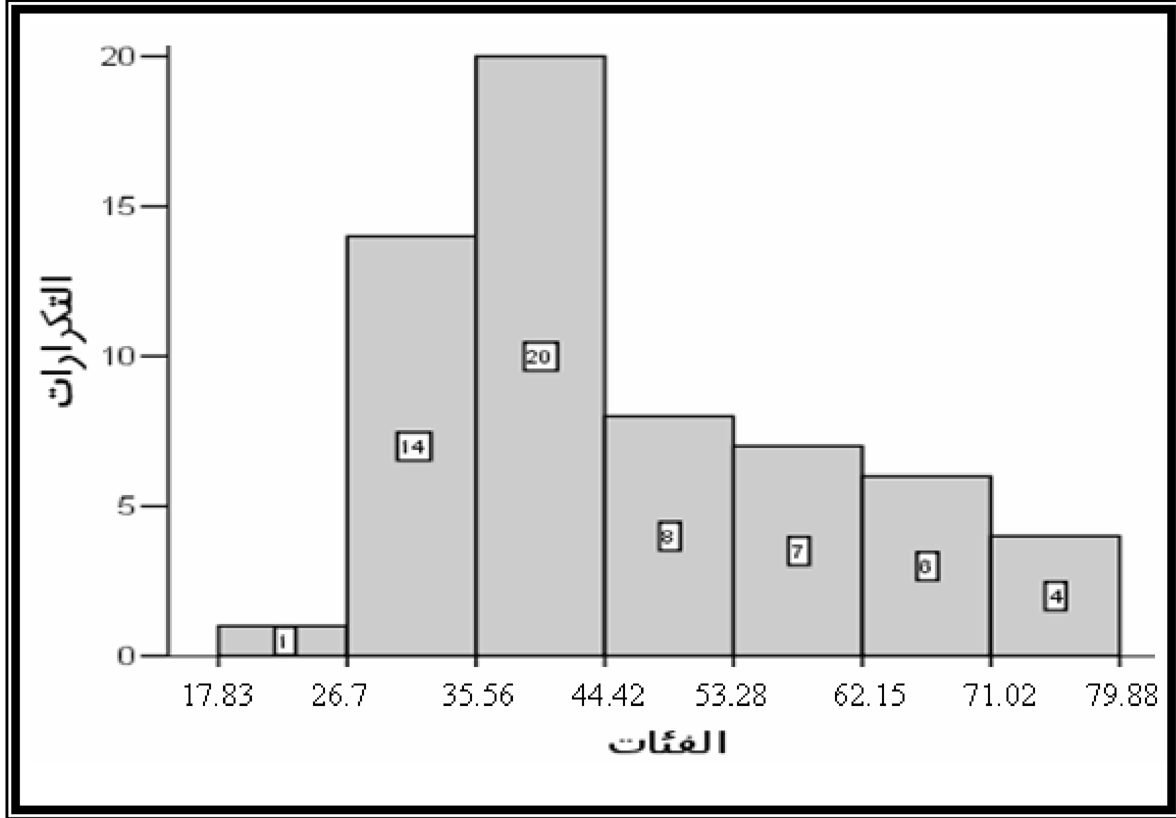
الشكل (1.2)

المدرج التكراري إلى 45 عامل حسب أجورهم اليومية

مثال (8.2) استخدم المدرج التكراري في تمثيل البيانات المبوبة في المثال (2.2):

تسلسل الفئة	الفئات (كمية إنتاج الحنطة بالطن)	التكرار f_i (عدد قطع الأراضي)	مراكز الفئات
1	26.70-17.83	1	22.27
2	35.56-26.70	14	31.13
3	44.42-35.56	20	39.00
4	53.28-44.42	8	48.86
5	62.15-53.28	7	57.73
6	71.02-62.15	6	66.59
7	79.88-71.02	4	75.46

الحل: بما أن البيانات من النوع المستمر لذلك فإن المدرج التكراري يتكون من مستطيلات متلاصقة، كما يوضحه الشكل الآتي:



الشكل (2.2)

المدرج التكراري إلى 60 قطعة زراعية حسب إنتاج الحنطة

ملاحظة: أحياناً تكون فئات التوزيع غير متساوية في الأطوال لذا يستوجب إجراء تعديلات في التكرارات من خلال قسمة التكرار الأصلي على طول الفئة لنحصل على تكرار معدل f^* ، أي أن:

$$f_i^* = \frac{f_i}{L_i} \quad ; \quad i = 1,2,\cdots,m \qquad \cdots \quad (2.11)$$

2- **المضلع التكراري** Frequency Polygon:

هو عبارة عن رسم بياني باستخدام سلسلة من المستقيمات التي تصل النقاط المتمثلة بالفئات مقابل التكرارات أو التكرارات النسبية وذلك من خلال إيجاد مراكز

الفئات التي تمثل المحور الأفقي وتحديد النقطة التي تقابل مركز كل فئة على المحور العمودي ومن ثم وصل مستقيمات بين النقاط التي حددت بعضها بالبعض.

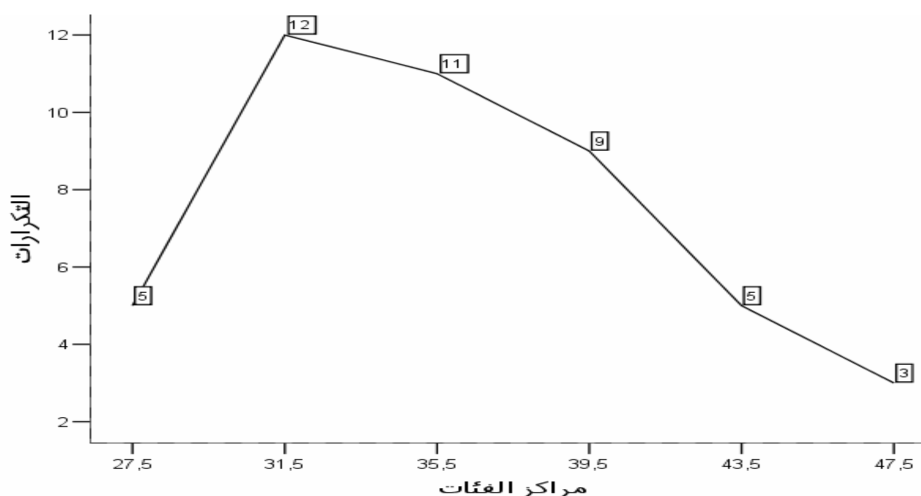
ويمكن رسم المضلع التكراري من خلال المدرج التكراري وذلك بتنصيف القواعد العليا لمستطيلات المدرج التكراري ثم نصل بين هذه النقاط بمستقيمات فنحصل على المضلع التكراري. كما يوضحه المثال الآتي:

مثال (9.2)

بالإعتماد على المثال (1.2) **أرسم المضلع التكراري**

تسلسل الفئة	الفئات (الأجور)	التكرار f_i (عدد العمال)	مراكز الفئات
1	29-26	5	27.5
2	33-30	12	31.5
3	37-34	11	35.5
4	41-38	9	39.5
5	45-42	5	43.5
6	49-46	3	47.5

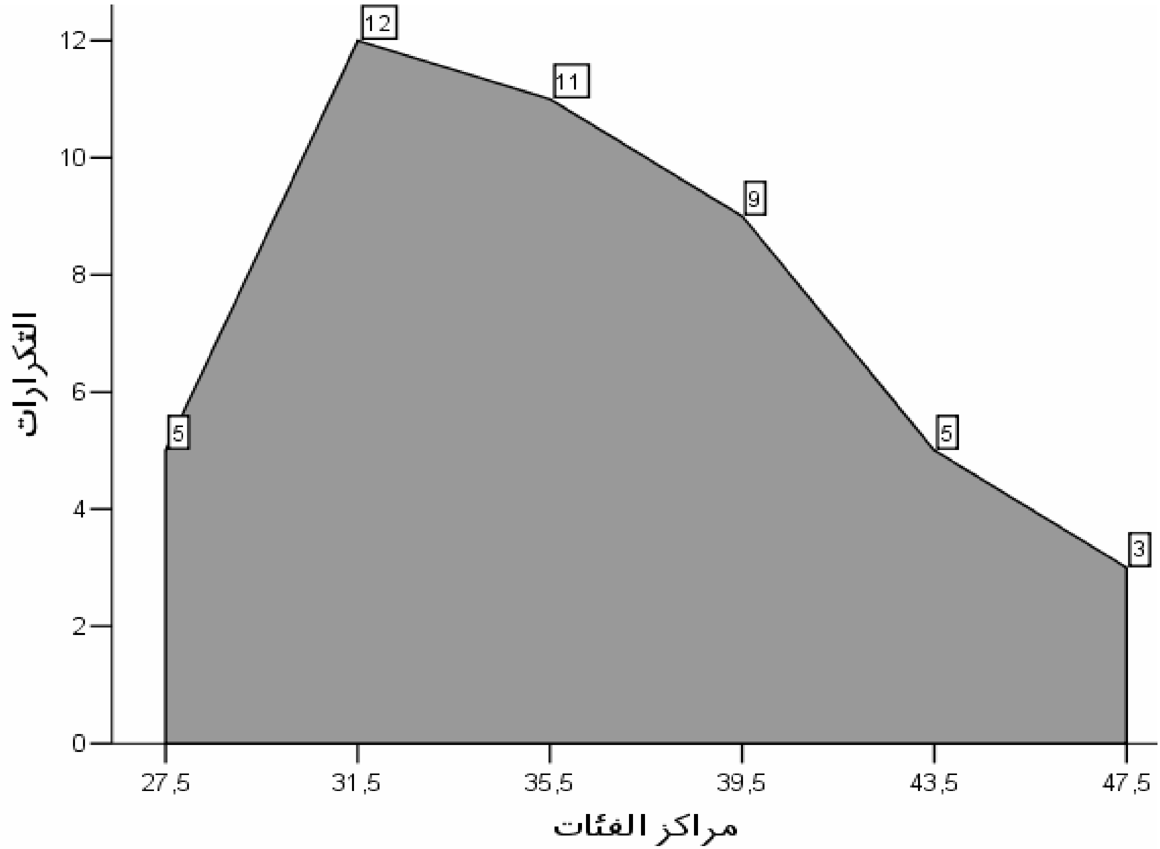
الحل:



الشكل (3.2)

المضلع التكراري إلى 45 عامل حسب أجورهم اليومية

ويمكن عرض المضلع التكراري في المثال السابق على شكل مساحة تمثل حجم التكرارات وكما يأتي:



الشكل (4.2)

المضلع التكراري (على شكل مساحة) إلى 45 عامل حسب أجورهم اليومية

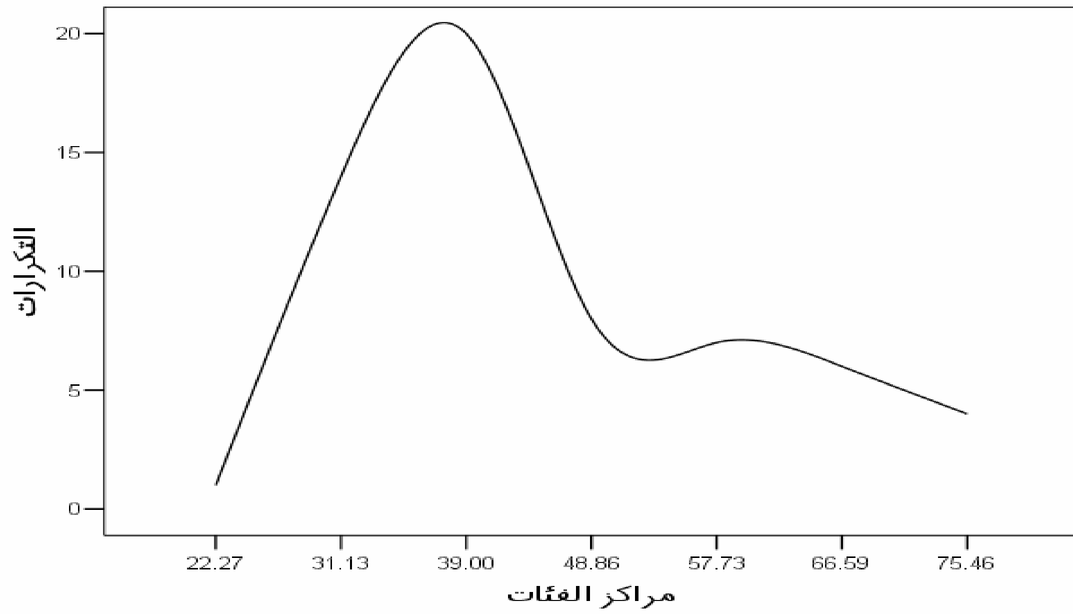
3- **المنحنى التكراري** Frequency Curve:

وهي طريقة مشابه لطريقة رسم المضلع التكراري ولكن بدلاً من أن نصل بين النقاط بمستقيمات نصل منحنياً مستمراً يمر بجميع هذه النقاط ويستخدم عادة مع البيانات المستمرة ويمكن توضيح المنحنى التكراري من خلال المثال الآتي:

مثال (10.2)**: بالإعتماد على المثال** (2.2) **أرسم المنحنى التكراري:**

تسلسل الفئة	الفئات (كمية إنتاج الحنطة بالطن)	التكرار f_i (عدد قطع الأراضي)	مراكز الفئات
1	26.70-17.83	1	22.27
2	35.56-26.70	14	31.13
3	44.42-35.56	20	39.00
4	53.28-44.42	8	48.86
5	62.15-53.28	7	57.73
6	71.02-62.15	6	66.59
7	79.88-71.02	4	75.46

الحل: يمكن رسم المنحنى التكراري للبيانات المستمرة كما في الشكل الآتي:



الشكل (5.2)

المنحنى التكراري إلى 60 قطعة زراعية حسب إنتاج الحنطة

4- **المنحنى التكراري المتجمع** Cumulative Frequency Curve :

لغرض رسم المنحنى التكراري المتجمع الصاعد يجب أولاً تكوين جدول تكراري متجمع صاعد ثم يتم تحديد النقاط على المحور الأفقي بالحدود العليا للفئات وتحدد النقاط على المحور العمودي بالتكرار المتجمع الصاعد (أو التكرار المتجمع الصاعد النسبي).

مثال (11.2)

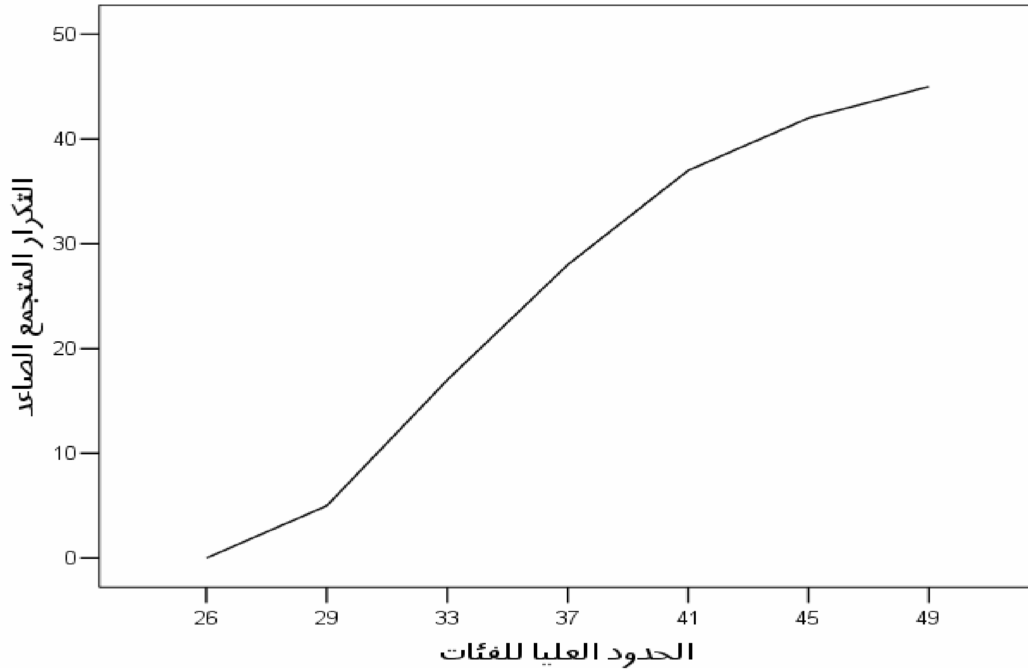
بالإعتماد على المثال (1.2) **أرسم المنحنى التكراري المتجمع الصاعد (والنسبي) أي للجدول التكراري الآتي:**

تسلسل الفئة	الفئات (الأجور)	التكرار f_i (عدد العمال)	التكرار المتجمع الصاعد F_i	التكرار المتجمع الصاعد النسبي $F_i^*\%$
1	29-26	5	5	11.111
2	33-30	12	17	37.778
3	37-34	11	28	62.222
4	41-38	9	37	82.222
5	45-42	5	42	93.333
6	49-46	3	45	100

الحل: لدينا الحدود العليا للفئات التي تمثل المحور الأفقي مقابل التكرار المتجمع الصاعد الذي يمثل المحور العمودي، أي أن:

تسلسل الفئة	الحدود العليا للفئات	التكرار المتجمع الصاعد F_i
1	29	5
2	33	17
3	37	28
4	41	37
5	45	42
6	49	45

يفضل غلق المنحنى بإضافة الحد الأدنى للفئة الأولى مقابل الصفر والشكل التالي يوضح المنحنى التكراري المتجمع الصاعد:



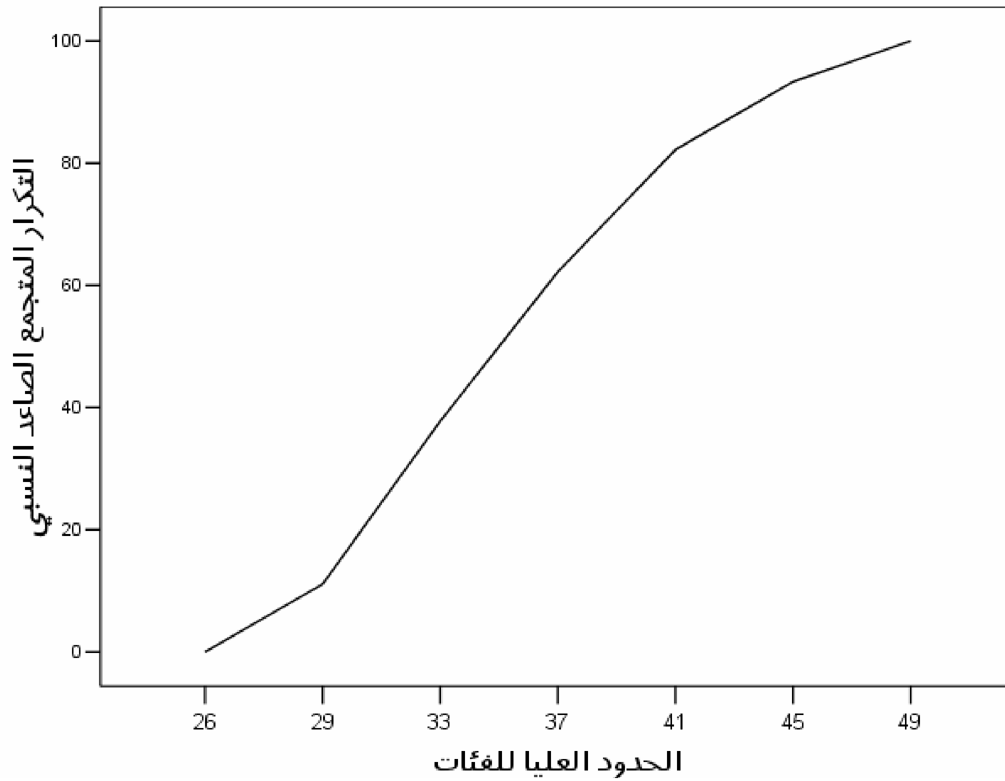
الشكل (6.2)

المنحنى التكراري المتجمع الصاعد إلى 45 عامل حسب أجورهم اليومية

وبنفس الطريقة أعلاه يمكن رسم المنحنى التكراري المتجمع الصاعد النسبي، أي أن:

تسلسل الفئة	الحدود العليا للفئات	التكرار المتجمع الصاعد النسبي $F_i^*\%$
1	29	11.111
2	33	37.778
3	37	62.222
4	41	82.222
5	45	93.333
6	49	100

كما يوضحه الشكل الآتي:



الشكل (7.2)

المنحنى التكراري المتجمع الصاعد النسبي

إلى 45 عامل حسب أجورهم اليومية

لغرض تمثيل المنحنى التكراري المتجمع النازل (والنسبي) يتم تحديد الحدود الدنيا للفئات، وبعد ذلك يتم تعيين النقاط التي إحداثياتها تمثل أزواج القيم (الحدود الدنيا للفئات، التكرار المتجمع النازل)، ويفضل غلق المنحنى بإضافة الحد الأعلى للفئة الأخيرة مقابل الصفر والمثال الآتي يوضح ذلك:

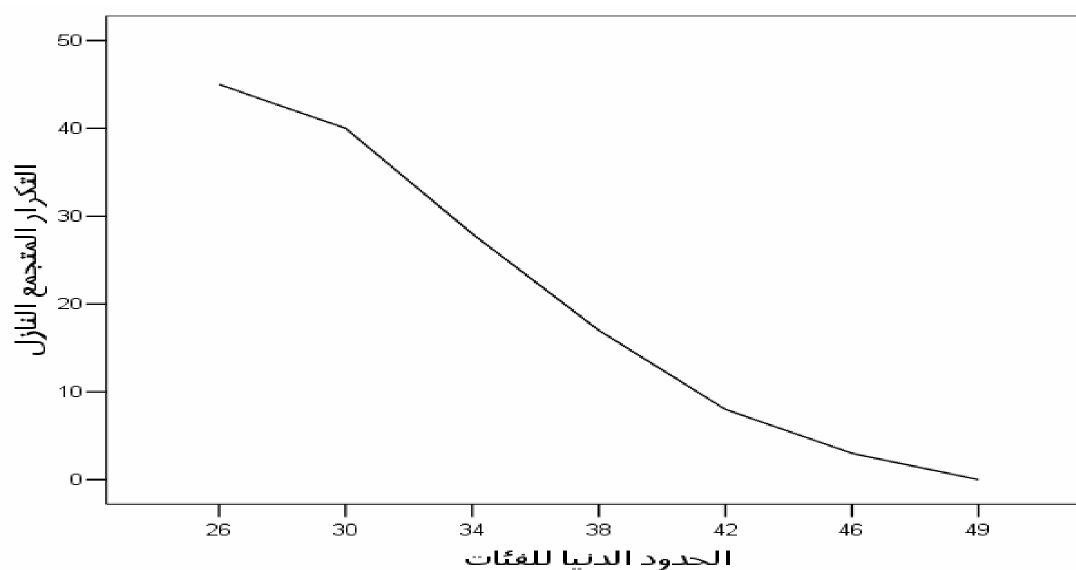
مثال (12.2): بالإعتماد على المثال (1.2) أرسم المنحنى التكراري المتجمع النازل (والنسبي)، أي للجدول التكراري الآتي:

تسلسل الفئة	الفئات (الأجور)	التكرار f_i (عدد العمال)	التكرار المتجمع النازل F'_i	التكرار المتجمع النازل النسبي F'^{*}_i %
1	29-26	5	45	100
2	33-30	12	40	88.889
3	37-34	11	28	62.222
4	41-38	9	17	37.778
5	45-42	5	8	17.778
6	49-46	3	3	6.667

الحل: يتم أولاً تحديد الحدود الدنيا للفئات مع الحد الأعلى للفئة الأخيرة مقابل التكرار المتجمع النازل F'_i، أي أن:

تسلسل الفئة	الفئات (الأجور)	التكرار المتجمع النازل F'_i
1	26	45
2	30	40
3	34	28
4	38	17
5	42	8
6	49-46	3

والشكل الآتي يوضح المنحنى التكراري المتجمع النازل:



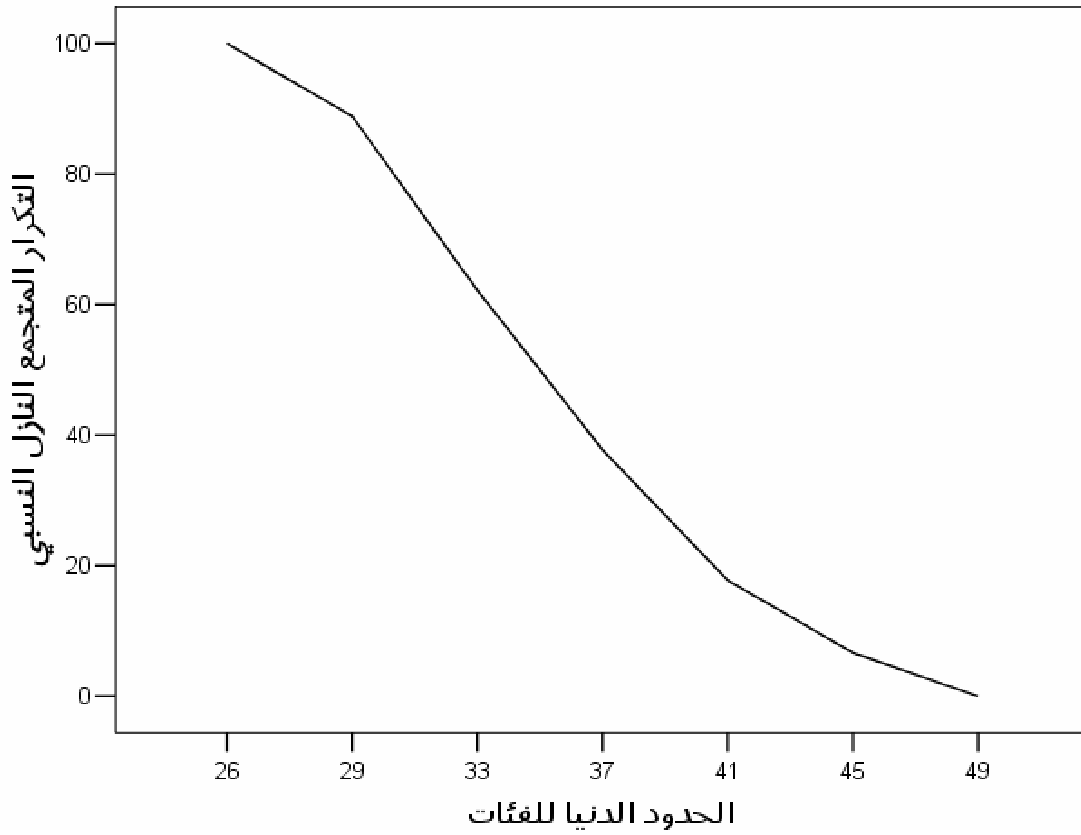
الشكل (8.2)

المنحنى التكراري المتجمع النازل إلى 45 عامل حسب أجورهم اليومية

لرسم المنحنى التكراري المتجمع النازل النسبي يجب تحديد الحدود الدنيا للفئات مع التكرار المتجمع النازل النسبي، أي أن:

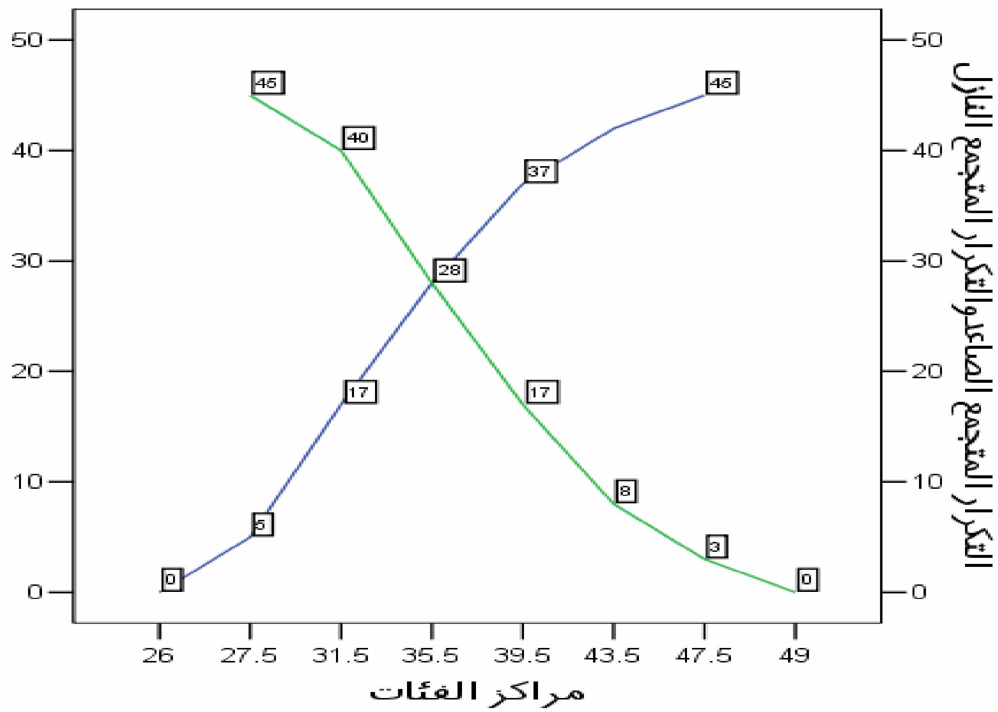
تسلسل الفئة	الفئات (الأجور)	التكرار المتجمع النازل النسبي $F_i'^*\%$
1	26	100
2	30	88.889
3	34	62.222
4	38	37.778
5	42	17.778
6	49-46	6.667

والشكل الآتي يوضح ذلك:



الشكل (9.2): المنحنى التكراري المتجمع النازل النسبي
إلى 45 عامل حسب أجورهم اليومية

ملاحظة: أحياناً يطلب رسم المنحنى التكراري المتجمع الصاعد والنازل مقابل مراكز الفئات في شكل موحد كما يوضحه الشكل التالي بالأعتماد على بيانات المثال (1.2):



الشكل (10.2)

المنحنى التكراري المتجمع الصاعد والنازل إلى 45 عامل حسب أجورهم اليومية

ثانياًّ: في حالة البيانات غير المبوبة لدينا ما يلي:

1- الأعمدة البيانية Bar Charts:

وتستخدم عادة في عرض البيانات المنفصلة وكذلك البيانات الوصفية وتوجد عدة أشكال مختلفة، منها ما يلي:

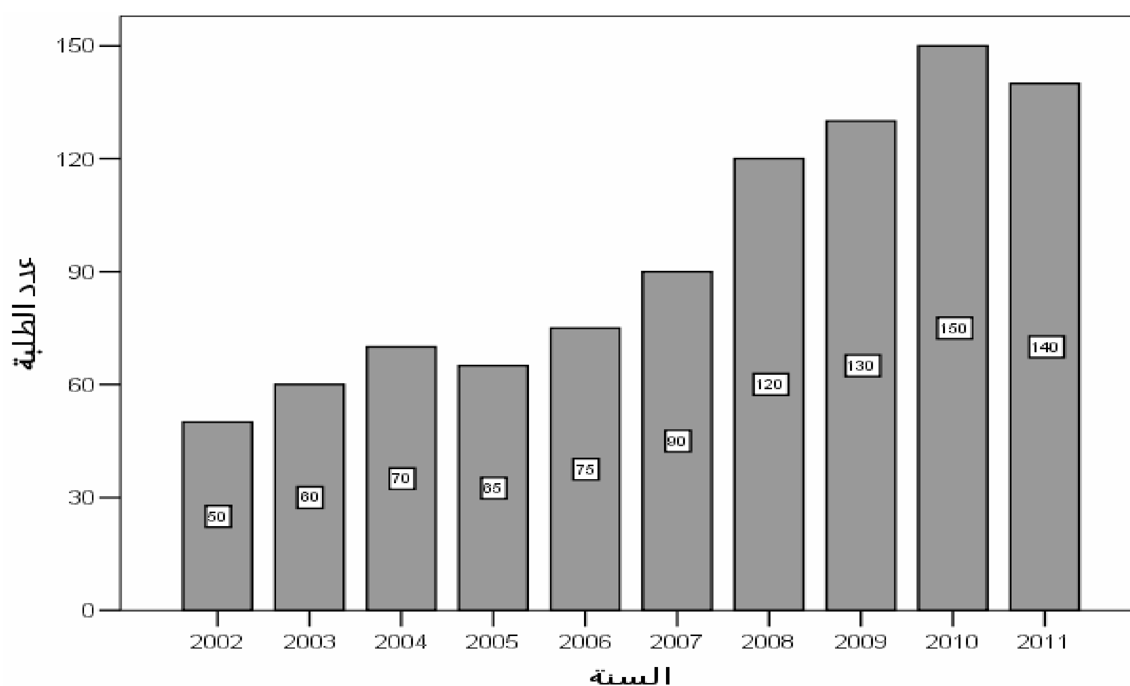
أ- **الأعمدة البيانية البسيطة Simple Bar Chart:** تستخدم عادة في تمثيل قيم البيانات لظاهرة واحدة المراد دراستها وقد تكون هذه البيانات مقاسة مقابل الزمن (سلسلة زمنية) أو تسلسل وحدات الإنتاج أو غير ذلك.

مثال (13.2)

استخدم الأعمدة البيانية في تمثيل عدد الطلبة المقبولين في قسم الإقتصاد خلال (10) سنوات.

السنة	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
عدد الطلبة	50	60	70	65	75	90	120	130	150	140

الحل: يمكن استخدام الأعمدة البيانية في تمثيل بيانات الجدول مـن خـلال فـرض أن الـسنة تمثـل المحـور الأفقي في حين يمثل أعداد الطلبة المقبولين في قسم الإقتصاد المحور العمودي والـشكل الآتي يوضـح ذلك:



الشكل (11.2)

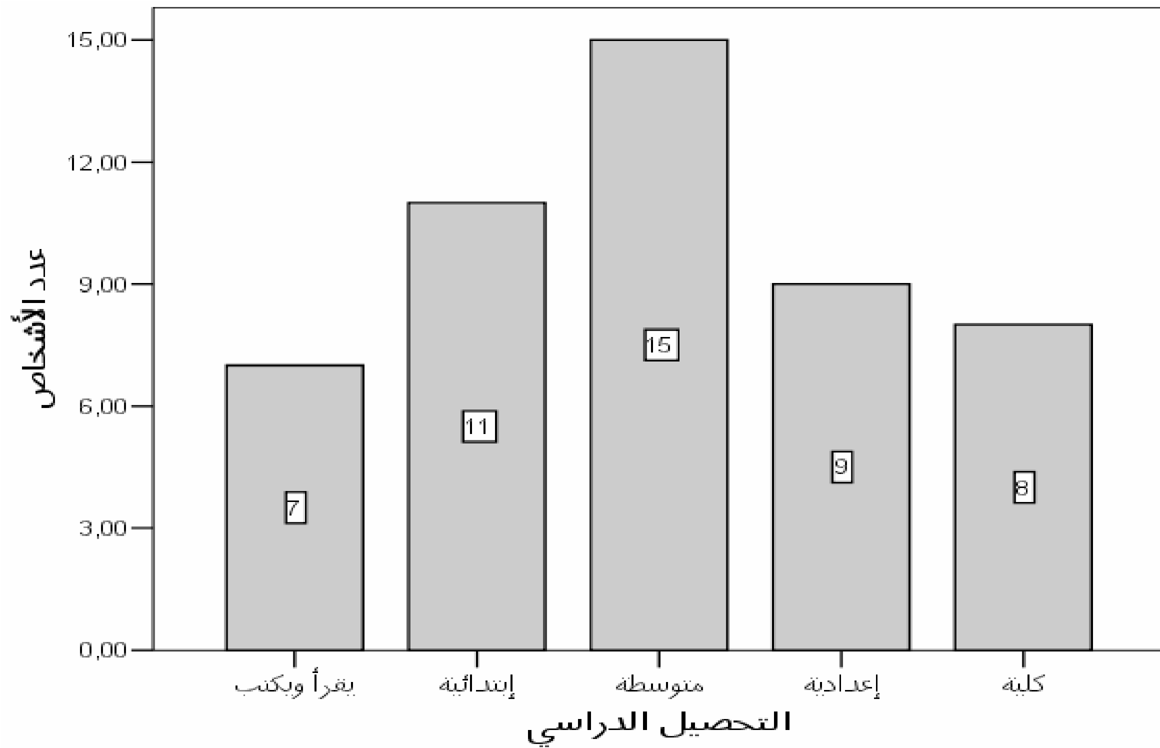
الأعمدة البيانية لأعداد الطلبة المقبولين في قسم الإقتصاد

خلال الفترة (2002-2011)

مثال (14.2)**: استخدم الأعمدة البيانية في تمثيل التحصيل العلمي لبيانات المثال** (6.2)**:**

تحصيلهم العلمي	عدد الأشخاص
يقرأ ويكتب	7
إبتدائية	11
متوسطة	15
إعدادية	9
كلية	8

الحل: التحصيل العلمي يمثل المحور الأفقي في حين يمثل أعداد الأشخاص اللذين شملتهم الدراسة المحور العمودي، كما يوضحه الشكل الآتي:



الشكل (12.2)

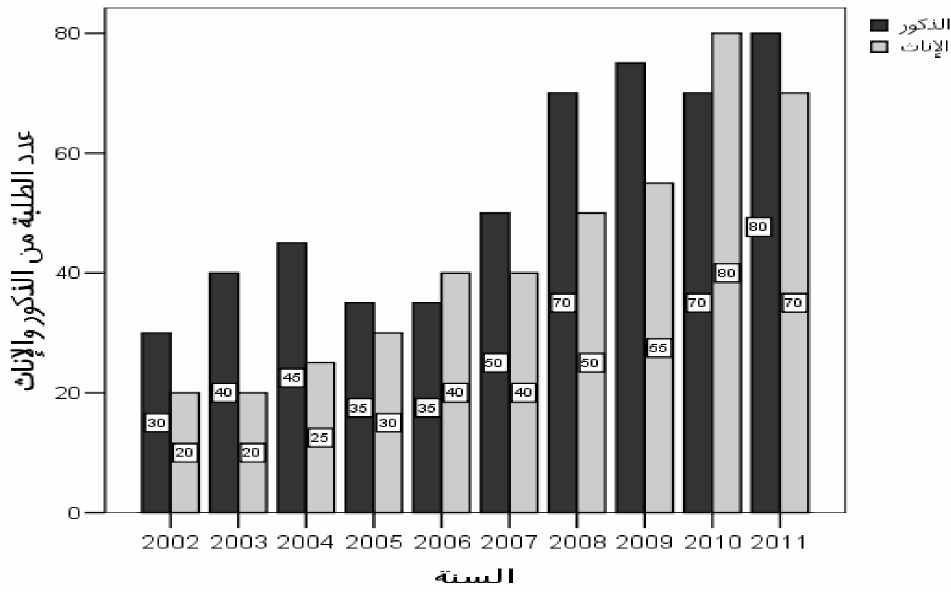
الأعمدة البيانية لأعداد الأشخاص حسب تحصيلهم العلمي

ب- **الأعمدة البيانية المركبة** Clustered or Stacked Bar Chart : تستخدم عادة في تمثيل قيم البيانـات لأكـثر من ظاهرة واحدة تحت الدراسة (ظاهرتين أو أكثر).

مثال (15.2)**: استخدم الأعمدة البيانية المركبة في تمثيل أعداد الطلبة المقبولين في قسم الإقتـصاد المـصنفين ذكور وإناث حسب السنوات (2002-2011) كما في الجدول الآتي:**

السنة	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
الذكور	30	40	45	35	35	50	70	75	70	80
الإناث	20	20	25	30	40	40	50	55	80	70

الحل: لرسم الأعمدة البيانية المركبة يمكن إعتبـار المحـور الأفقـي يمثـل الـسنوات في حـين يمثـل المحـور العمودي عدد الطلبة من الذكور والإناث وكما يوضحه الشكل الآتي:

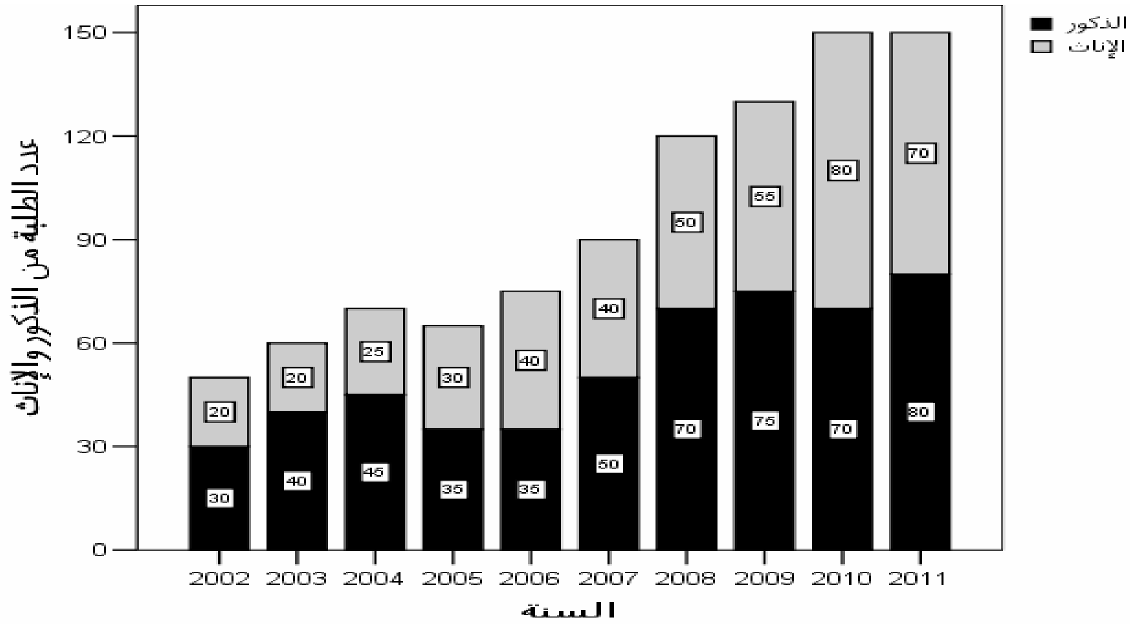


الشكل (13.2)

الأعمدة البيانية المركبة لعدد الطلبة المقبولين في قسم الإقتصاد من الذكور والإناث خلال الفترة (2002-2011)

ويمكــن تمثيــل البيانــات عــلى شــكل أعمــدة مجــزأة إلى ذكــور وإنــاث، كــما يوضــحه

الشكل الآتي:



الشكل (14.2)

الأعمدة البيانية المركبة (مجزأة) لعدد الطلبة المقبولين في قسم الإقتصاد من الذكور والإناث خلال الفترة (2002-2011)

2- الدائرة البيانية Pie Chart:

تستخدم عادةً مع البيانات الوصفية مثل (أنواع القطاعات الإقتصادية، التوزيع السكاني، أنواع السلع الإستهلاكية،...الخ). حيث تقسم الدائرة إلى عدة أقسام عددها يساوي عدد القطاعات المدروسة والتي يتم تمثيلها داخل الدائرة بحيث أن مجموع مساحات القطاعات تمثل مساحة الدائرة وبهدف تحديد حجم كل قطاع فإنه يستوجب تحديد زاوية كل منها، أي أن زاوية القطاع θ_i هي كما يلي:

$$\theta_i = \frac{A_i}{T} \times 360^\circ \qquad \cdots \quad (2.12)$$

A_i : تمثل عدد بيانات النوع.

T: تمثل مجموع البيانات الكلية.

مثال (16.2)**: تناول أحد البحوث الإقتصادية إنفاق الأسرة الشهري البالغ 900 ألف دينار موزعة على النحو الآتي: الغذاء 300، الملابس 200، المسكن 250، الكهرباء 100، متفرقة 50. والمطلوب رسم الدائرة البيانية لتمثيل هذه البيانات.**

الحل:

بإستخدام الصيغة (2.12) يمكن الحصول على كل زاوية من زوايا أنواع الإنفاق وكما يلي:

$$\theta_i = \frac{A_i}{T} \times 360^\circ$$

$$\theta_1 = \frac{A_1}{T} \times 360^\circ = \frac{300}{900} \times 360^\circ = 120^\circ$$ الغذاء

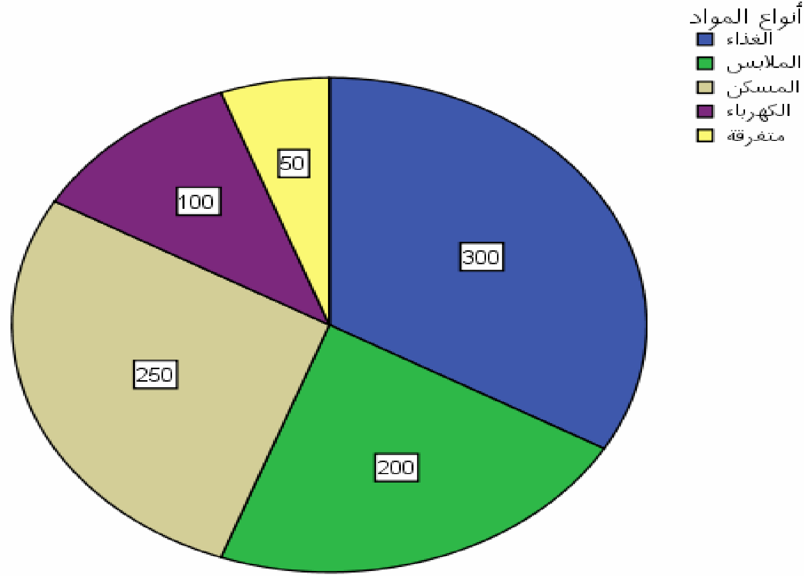
$$\theta_2 = \frac{A_2}{T} \times 360^\circ = \frac{200}{900} \times 360^\circ = 80^\circ$$ الملابس

$$\theta_3 = \frac{A_3}{T} \times 360^\circ = \frac{250}{900} \times 360^\circ = 100^\circ$$ المسكن

$$\theta_4 = \frac{A_4}{T} \times 360^\circ = \frac{100}{900} \times 360^\circ = 40^\circ$$ الكهرباء

$$\theta_5 = \frac{A_5}{T} \times 360^\circ = \frac{50}{900} \times 360^\circ = 20^\circ$$ متفرقة

بالإعتماد على الزوايا المحسوبة أعلاه يمكن رسم الدائرة بالشكل الآتي:



الشكل (15.2)

الدائرة البيانية لإنفاق الأسرة الشهري في إقليم كوردستان

3- الخط البياني Line Chart:

يمكن تمثيل ظاهرة أو عدة ظواهر تحت الدراسة خلال فترة زمنية بشكل خط بياني والمثال التالي يوضح ذلك:

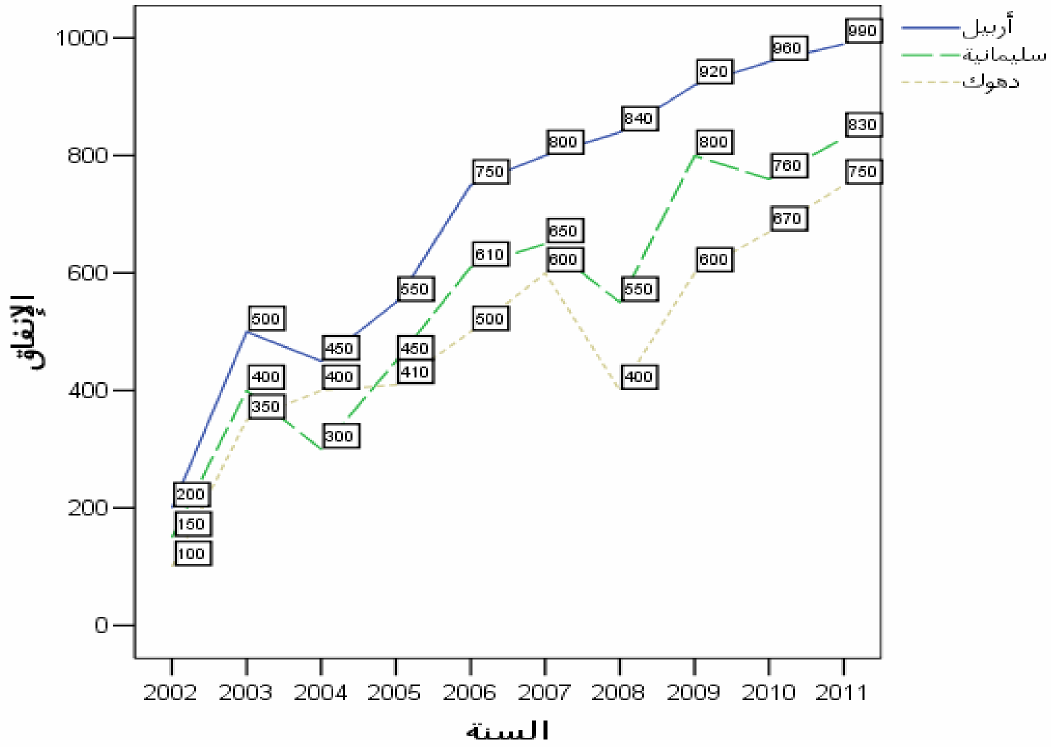
مثال (17.2)**: الجدول الآتي يمثل النفقات المفترضة لبلديات إقليم كوردستان خلال 10 سنوات:**

السنة	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
أربيل	200	500	450	550	750	800	840	920	960	990
سليمانية	150	400	300	450	610	650	550	800	760	830
دهوك	100	350	400	410	500	600	400	600	670	750

والمطلوب رسم الخط البياني الذي يمثل النفقات المفترضة لبلديات أقليم كوردستان خلال 10 سنوات.

الحل:

يمكن رسم الخط البياني للنفقات المفترضة لبلديات أقليم كوردستان التي تمثل المحور العمـودي مقابل السنوات التي تمثل المحور الأفقي، كما يوضحه الشكل الآتي:



الشكل (16.2)

الخط البياني لإنفاق الأسرة الشهري في إقليم كوردستان

تمارين الفصل الثاني

2-1: تم دراسة أوزان 50 تلميذ وكانت كما يلي:

27	43	33	37	35	24	23	36	37	19
43	24	23	42	37	33	24	27	43	18
19	33	28	45	23	22	34	30	32	23
42	22	31	33	34	44	31	31	41	34
32	19	35	45	29	45	32	22	34	33

المطلوب: نظم هذه البيانات في جدول تكراري وأحسب مراكز الفئة.

2-2: إذا توفرت لديك البيانات الأتية:

12.7	5.8	4.3	13.7	8.5	9.1	7.2	3.8	3.5	3.9
6.4	2.4	2.3	4.2	3.7	3.3	2.3	9.7	4.3	5.8
3.9	3.3	2.8	4.5	2.3	5.2	3.1	3.0	3.2	6.3
4.2	5.2	3.1	3.3	5.4	4.4	3.9	3.1	4.7	8.4

جد ما يلي:

1- **جدول التوزيع التكراري.**
2- **التكرار المتجمع الصاعد.**
3- **التكرار المتجمع الصاعد النسبي.**
4- **التكرار المتجمع النازل.**
5- **التكرار المتجمع النازل النسبي.**

2-3: فيما يلي عينة عشوائية تتكون من 100 طالب وتقديراتهم في مادة الإقتصاد:

التقدير في مادة الإقتصاد	عدد الطلاب
ضعيف	15
مقبول	22
متوسط	25
جيد	17
جيد جداً	12
إمتياز	9

المطلوب تكوين جدول التوزيع التكراري الصاعد والنازل؟

2-4: بالإعتماد على السؤال (2-1) **أرسم المدرج التكراري و المضلع التكراري.**

2-5: بالإعتماد على السؤال (2-2) **أرسم المدرج التكراري و المنحنى التكراري.**

2-6: بالإعتماد على السؤال (2-1) **أرسم المحنى التكراري المتجمع الصاعد والنسبي.**

2-7: بالإعتماد على السؤال (2-2) **أرسم المحنى التكراري المتجمع النازل والنسبي.**

2-8: إذا كانت لديك البيانات التي تمثل إنتاج الحنطة والشعير في إقليم كوردستان بألآف الأطنــــــان خلال خمسة سنوات كما في الجدول الآتي:

السنة	2006	2007	2008	2009	2010
إنتاج الحنطة	120	180	210	190	230
إنتاج الشعير	100	150	170	210	200

والمطلوب ما يلي:

1- استخدم الأعمدة البيانية في تمثيل إنتاج الحنطة في إقليم كوردستان خلال خمس سنوات.

2- استخدم الأعمدة البيانية في تمثيل إنتاج الحنطة والشعير في إقليم كوردستان خلال خمس سنوات.

2-9: أرسم الدائرة البيانية للإستهلاك اليومي من البنزين في محافظات إقليم كوردستان:

المحافظة	إستهلاك البنزين
أربيل	800
سليمانية	600
دهوك	450

2-10: أرسم الخط البياني للإيرادات والنفقات لشركة شيشار خلال 6 سنوات، وكما يلي:

السنة	2005	2006	2007	2008	209	2010
إيرادات	1200	2180	2210	2190	2300	3000
نفقات	1000	1150	1301	1100	1200	1500

الفصل الثالث

رموز ومصطلحات رياضية

Notations

الفصل الثالث

رموز ومصطلحات رياضية

Notations

1.3: مقدمة:

قبل البدء بدراسة الإحصاء الكمي وقياس بعض المؤشرات الإحصائية للبيانات المتاحة سوف نستعرض في هذا الفصل أهم الرموز والمصطلحات المستخدمة لاحقاً في الفصول القادمة، وهي كما يلي:

2.3: رمز الجمع $\sum$:

أفترض أن x تعني درجات مادة الإقتصاد لعينة عشوائية (n=10) تمثل طلبة قسم الإقتصاد، حيث أن x_1 تمثل درجة مادة الإقتصاد للطالب الأول، x_2 تمثل درجة مادة الإقتصاد للطالب الثاني،...، إلى x_{10} تمثل درجة مادة الإقتصاد للطالب الأخير، ويلاحظ أننا حصلنا على نتائج درجات الطلبة التي تشكل سلسلة من الأعداد هي:

$$x_1 , x_1 , \cdots \ x_{10}$$

وعندئذٍ فإن المجموع الكلي لهذه الأعداد هو: $x_1 + x_1 + \cdots + x_{10}$ ولتسهيل عملية كتابة هذا المجموع بشكل مختصر يتم التعبير عنه بالشكل: $\sum_{i=1}^{10} x_i$ حيث أن الرمز $\sum$ يشير إلى عملية جمع، وهو حرف إغريقي يلفظ Sigma، وأن i تمثل دليل (مؤشر) Subscript لتسلسل العدد عند عملية الجمع فإذا كانت i =1 فذلك يعني درجة الطالب الأول، i =2 تعني درجة الطالب الثاني، وهكذا البقية إلى i =10 تعني درجة الطالب الأخير لهذه العينة. والعملية $\sum_{i=1}^{10} x_i$ تقرأ إجمالاً على النحو الآتي: مجموع سلسلة من الأعداد إبتداءاً بدرجة الطالب الأول i =1 وإنتهاءاً بدرجة الطالب الأخير i =10 وبشكل عام إذا كان

لدينا سلسلة من الكميات عددها n فإن المجموع الكلي لها يتم التعبير عنه بالـشكل $\sum_{i=1}^{n} x_i$ ، وعـلى هـذا الأساس يمكن التعبير عن سلسلة كميات هـي $x_1, x_1, \cdots x_n$ لعـدة عمليـات مختلفـة بالـشكل الآتي:

العملية المطلوبة	العملية بشكل رموز	رمز الجمع
مجموع قيم بيانات السلسلة	$x_1 + x_2 + \cdots + x_n$	$\sum_{i=1}^{n} x_i$
مجموع مربعات قيم بيانات السلسلة	$x_1^2 + x_2^2 + \cdots + x_n^2$	$\sum_{i=1}^{n} x_i^2$
مربع مجموع قيم بيانات السلسلة	$(x_1 + x_1 + \cdots + x_n)^2$	$\left(\sum_{i=1}^{n} x_i\right)^2$
مجموع مقلوب قيم بيانات السلسلة	$\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}$	$\sum_{i=1}^{n}\left(\frac{1}{x_i}\right)$
مجموع مربعات مقلوب بيانات السلسلة	$\frac{1}{x_1^2} + \frac{1}{x_2^2} + \cdots + \frac{1}{x_n^2}$	$\sum_{i=1}^{n}\left(\frac{1}{x_i^2}\right) = \sum_{i=1}^{n}\left(\frac{1}{x_i}\right)^2$
مقلوب مجموع قيم بيانات السلسلة	$\frac{1}{x_1 + x_2 + \cdots + x_n}$	$\frac{1}{\sum_{i=1}^{n} x_i}$
مقلوب مجموع مربعات قيم بيانات	$\frac{1}{x_1^2 + x_2^2 + \cdots + x_n^2}$	$\frac{1}{\sum_{i=1}^{n} x_i^2}$

رمز الجمع	العملية بشكل رموز	العملية المطلوبة
		السلسلة
$\sum_{i=1}^{n} \log x_i$	$\log x_1 + \log x_2 + \cdots + \log x_n$	مجموع لوغارتمات قيم بيانات السلسلة
$\log \sum_{i=1}^{n} x_i$	$\log(x_1 + x_2 + \cdots + x_n)$	لوغارتم مجموع قيم بيانات السلسلة
$\sum_{i=1}^{n} \sqrt{x_i}$	$\sqrt{x_1} + \sqrt{x_2} + \cdots + \sqrt{x_n}$	مجموع جذور قيم بيانات السلسلة
$\sqrt{\sum_{i=1}^{n} x_i}$	$\sqrt{x_1 + x_2 + \cdots + x_n}$	جذر مجموع قيم بيانات السلسلة

أفرض أن هنالك سلسلتين من البيانات x و y عدد كل واحد منهما n من القيم ولدينا a و b تمثلان قيمتان ثابتتان، فإن:

مجموع الكمية الثابتة a إلى n من المرات هي:

$$\sum_{i=1}^{n} a = a\sum_{i=1}^{n}(1) = a \cdot n$$

1- مجموع حاصل ضرب الثابت a بقيم بيانات السلسلة x_i هي:

$$\sum_{i=1}^{n} a x_i = a \sum_{i=1}^{n} x_i$$

2- مجموع حاصل جمع أو طرح الثابت a إلى قيم بيانات السلسلة x_i هي:

$$\sum_{i=1}^{n}\left(x_i \mp a\right)=\sum_{i=1}^{n} x_i \mp \sum_{i=1}^{n} a=\sum_{i=1}^{n} x_i \mp a \sum_{i=1}^{n}(1)=\sum_{i=1}^{n} x_i \mp a \cdot n$$

3- مجموع حاصل جمع أو طرح بين قيم بيانات السلسلة x_i و قيم بيانات السلسلة y_i هـي:

$$\sum_{i=1}^{n}\left(x_i \mp y_i\right)=\sum_{i=1}^{n} x_i \mp \sum_{i=1}^{n} y_i$$

4- مجموع حاصل ضرب قيم بيانات السلسلة x_i بقيم بيانات السلـسلة y_i المقابلـة لها هـو:

$$\sum_{i=1}^{n} x_i y_i=x_1 y_1+x_2 y_2+\cdots+x_n y_n$$

5- مجموع حاصل ضرب مجموعتين تمثل المجموعـة الأولى قـيم بيانـات السلـسلة x_i مـضافاً (أو مطروح) لها القيمة الثابتة a مع المجموعة الثانية التي تمثل قيم بيانات السلـسلة y_i مـضافاً (أو مطروح) لها القيمة الثابتة b وهي:

$$\sum_{i=1}^{n}\left(x_i+a\right)\left(y_i+b\right)=\sum_{i=1}^{n}\left(x_i y_i+b\, x_i+a\, y_i+ab\right)$$

$$=\sum_{i=1}^{n} x_i y_i+b \sum_{i=1}^{n} x_i+a \sum_{i=1}^{n} y_i+abn$$

مثال (1.3)**: أفرض لديك البيانات الآتية:**

x_i	4	2	1	5
y_i	-3	-1	0	6

a = 8 و b = 10 جد ناتج مايلي:

7	6	5	4	3	2	1
$\frac{1}{\sum_{i=1}^{n} x_i^2}$	$\frac{1}{\sum_{i=1}^{n} x_i}$	$\sum_{i=1}^{n}\left(\frac{1}{x_i}\right)^2$	$\sum_{i=1}^{n}\left(\frac{1}{x_i}\right)$	$\left(\sum_{i=1}^{n} x_i\right)^2$	$\sum_{i=1}^{n} x_i^2$	$\sum_{i=1}^{n} x_i$
14	13	12	11	10	9	8
$\frac{ab}{\sum_{i=1}^{n} x_i}$	$\sum_{i=1}^{n} a x_i$	$\sum_{i=1}^{n} a$	$\sqrt{\sum_{i=1}^{n} x_i}$	$\sum_{i=1}^{n} \sqrt{x_i}$	$\log \sum_{i=1}^{n} x_i$	$\sum_{i=1}^{n} \log x_i$
20	19	18	17		16	15
$\sum_{i=1}^{n}(x_i - y_i)^2$	$\sum_{i=1}^{n}(x_i - y_i)$	$\sum_{i=1}^{n}(x_i + a)$	$\sum_{i=1}^{n}(x_i + a)(y_i + b)$		$\sum_{i=1}^{n} x_i y_i^2$	$\sum_{i=1}^{n} x_i y_i$

الحل: نلاحظ من خلال البيانات أعلاه أن n =4 ويمكن إيجاد ناتج التعبير الرياضي كما يلي:

1 $$\sum_{i=1}^{n} x_i = \sum_{i=1}^{4} x_i = x_1 + x_2 + x_3 + x_4 = 4 + 2 + 1 + 5 = 12$$

2 $$\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{4} x_i^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2 = 4^2 + 2^2 + 1^2 + 5^2 = 46$$

3 $$\left(\sum_{i=1}^{4} x_i\right)^2 = (x_1 + x_2 + x_3 + x_4)^2 = (4 + 2 + 1 + 5)^2 = 12^2 = 144$$

4 $$\sum_{i=1}^{4}\left(\frac{1}{x_i}\right) = \frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \frac{1}{x_4} = \frac{1}{4} + \frac{1}{2} + \frac{1}{1} + \frac{1}{5} = 0.25 + 0.5 + 1 + 0.2 = 1.95$$

5 $$\sum_{i=1}^{4}\left(\frac{1}{x_i}\right)^2=\sum_{i=1}^{4}\left(\frac{1}{x_i^2}\right)=\frac{1}{x_1^2}+\frac{1}{x_2^2}+\frac{1}{x_3^2}+\frac{1}{x_4^2}=\frac{1}{4^2}+\frac{1}{2^2}+\frac{1}{1^2}+\frac{1}{5^2}$$
$$=\frac{1}{16}+\frac{1}{4}+\frac{1}{1}+\frac{1}{25}=0.0625+0.25+1+0.04=1.3525$$

6 $$\frac{1}{\sum_{i=1}^{4} x_i}=\frac{1}{12}=0.0833$$

7 $$\frac{1}{\sum_{i=1}^{4} x_i^2}=\frac{1}{46}=0.0217$$

8 $$\sum_{i=1}^{4}\log x_i=\log x_1+\log x_2+\log x_3+\log x_4$$
$$=\log 4+\log 2+\log 1+\log 5$$
$$=0.6021+0.3010+0+0.699=1.602$$

9 $$\log\sum_{i=1}^{4} x_i=\log 12=1.0792$$

10 $$\sum_{i=1}^{4}\sqrt{x_i}=\sqrt{x_1}+\sqrt{x_2}+\sqrt{x_2}+\sqrt{x_4}=\sqrt{4}+\sqrt{2}+\sqrt{1}+\sqrt{5}$$
$$=2+1.4142+1+2.2361=6.6503$$

11 $$\sqrt{\sum_{i=1}^{4} x_i}=\sqrt{12}=3.4641$$

12 $$\sum_{i=1}^{4}\mathrm{a}=\mathrm{a}\sum_{i=1}^{4}(1)=\mathrm{a}\cdot 4=(8)(4)=32$$

13 $$\sum_{i=1}^{4} \mathrm{a}x_i = \mathrm{a}\sum_{i=1}^{4} x_i = (8)(12) = 96$$

14 $$\frac{\mathrm{ab}}{\sum_{i=1}^{4} x_i} = \frac{(8)(10)}{12} = 6.6667$$

15 $$\sum_{i=1}^{4} x_i y_i = x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4 = (4)(-3)+(2)(-1)+(1)(0)+(5)(6)$$
$$= -12-2+0+30 = 16$$

16 $$\sum_{i=1}^{4} x_i y_i^2 = x_1y_1^2 + x_2y_2^2 + x_3y_3^2 + x_4y_4^2$$
$$= (4)(-3)^2 + (2)(-1)^2 + (1)(0)^2 + (5)(6)^2 = 36+2+0+180 = 218$$

17 $$\sum_{i=1}^{4} (x_i + \mathrm{a})(y_i + \mathrm{b}) = \sum_{i=1}^{4} (x_i y_i + \mathrm{b}\,x_i + \mathrm{a}\,y_i + \mathrm{ab})$$
$$= \sum_{i=1}^{4} x_i y_i + \mathrm{b}\sum_{i=1}^{4} x_i + \mathrm{a}\sum_{i=1}^{4} y_i + \mathrm{abn}$$
$$= 16 + (10)(12) + (8)(2) + (8)(10)(4) = 472$$

18 $$\sum_{i=1}^{4} (x_i + \mathrm{a}) = \sum_{i=1}^{4} x_i + \sum_{i=1}^{4} \mathrm{a} = \sum_{i=1}^{4} x_i + \mathrm{a}\sum_{i=1}^{4} (1)$$
$$= \sum_{i=1}^{4} x_i + \mathrm{a}\cdot 4 = 12 + (8)(4) = 44$$

19 $$\sum_{i=1}^{4} (x_i - y_i) = \sum_{i=1}^{4} x_i - \sum_{i=1}^{4} y_i = 12 - 2 = 10$$

20 $$\sum_{i=1}^{n}(x_i - y_i)^2 = (x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2 + (x_4 - y_4)^2$$

$$= (4-(-3))^2 + (2-(-1))^2 + (1-0)^2 + (5-6)^2$$

$$= (7)^2 + (3)^2 + (1)^2 + (-1)^2 = 60$$

2.3: رمز الضرب $\prod$:

أحياناً نحتاج إلى عملية ضرب مجموعة من الكميات ببعضها مما يتطلب الأمر إلى ترميز العملية لتسهيل كتابتها والرمز هو $\prod$ الذي يدل على وجود عملية ضرب مجموعة من الكميات، فإذا كان لدينا مجموعة قيم بيانات سلسلة هي $x_1 , x_2 , \cdots x_n$ فإن حاصل ضرب هذه الكميات ببعضها هو كما يلي:

$$\prod_{i=1}^{n} x_i = x_1 . x_2 . \cdots . x_n$$

وتقرأ هذه العملية إجمالاً " حاصل ضرب قيم بيانات سلسلة من الكميات ببعضها إبتداءً بالقيمة الأولى $i = 1$، وإنتهاءً بالقيمة الأخيرة $i = n$. وإذا فرضنا أن هنالك سلسلتين من البيانات x و y عدد كل واحد منهما n من القيم ولدينا a و b تمثلان قيمتان ثابتتان، فإن:

رمز الضرب	مفكوك وتبسيط العملية
$\prod_{i=1}^{n} a$	a^n
$\prod_{i=1}^{n} a\, x_i$	$a^n \cdot \prod_{i=1}^{n} x_i$

$\prod_{i=1}^{n} a\,b\,x_i y_i$	$(ab)^n \prod_{i=1}^{n} x_i \prod_{i=1}^{n} y_i$
$\prod_{i=1}^{n} \frac{1}{x_i}$	$\frac{1}{\prod_{i=1}^{n} x_i}$
$\log \prod_{i=1}^{n} x_i$	$\sum_{i=1}^{n} \log x_i$
$\prod_{i=1}^{n} \sqrt{x_i}$	$\sqrt{\prod_{i=1}^{n} x_i}$
$\prod_{i=1}^{n} x_i^a$	$\left(\prod_{i=1}^{n} x_i\right)^a$
$\log \prod_{i=1}^{n} x_i^a$	$a \sum_{i=1}^{n} \log x_i$
$\prod_{i=1}^{n} \frac{x_i}{y_i}$	$\frac{\prod_{i=1}^{n} x_i}{\prod_{i=1}^{n} y_i}$

مثال (2.3)**: أفرض لديك البيانات الآتية:**

x_i	7	1	5
y_i	2	0	4

$a = 6$ و $b = 8$ **جد ناتج مايلي:**

5	4	3	2	1
$\prod_{i=1}^{n} a\,b\,x_i y_i$	$\prod_{i=1}^{n} a\,x_i$	$\prod_{i=1}^{n} a$	$\prod_{i=1}^{n} x_i^2$	$\prod_{i=1}^{n} x_i$
10	9	8	7	6
$\log \prod_{i=1}^{n} x_i^{a\,y_i}$	$\log \prod_{i=1}^{n} x_i^{a}$	$\log \prod_{i=1}^{n} x_i$	$\prod_{i=1}^{n} \frac{1}{x_i^2}$	$\prod_{i=1}^{n} \frac{1}{x_i}$
15	14	13	12	11
$\prod_{i=1}^{n} (x_i - a)^2$	$\prod_{i=1}^{n} (x_i - y_i)^2$	$\prod_{i=1}^{n} x_i y_i^2$	$\prod_{i=1}^{n} x_i^{a}$	$\prod_{i=1}^{n} \sqrt{x_i}$

الحل: نلاحظ من خلال البيانات أعلاه أن $n = 3$ ويمكن إيجاد ناتج التعبير الرياضي كما يلي:

1
$$\prod_{i=1}^{n} x_i = \prod_{i=1}^{3} x_i = (x_1 \cdot x_2 \cdot x_3) = (7)(1)(5) = 35$$

2
$$\prod_{i=1}^{3} x_i^2 = x_1^2 \cdot x_2^2 \cdot x_3^2 = (7)^2 (1)^2 (5)^2 = 1225$$

$$or \qquad = (x_1 \cdot x_2 \cdot x_3)^2 = [(7)(1)(5)]^2 = [35]^2 = 1225$$

3
$$\prod_{i=1}^{3} a = a^3 = 6^3 = 216$$

4
$$\prod_{i=1}^{3} a\,x_i = a^3 \cdot \prod_{i=1}^{3} x_i = 6^3 \cdot 35 = 7560$$

5 $$\prod_{i=1}^{3} ab\, x_i y_i = (ab)^3 \prod_{i=1}^{3} x_i \prod_{i=1}^{3} y_i = (6\cdot 8)^2\, 35\cdot 0$$

$$= (2304)(35)(0) = 0 \qquad \textit{where} \quad \prod_{i=1}^{3} y_i = (2)(0)(4) = 0$$

6 $$\prod_{i=1}^{3} \frac{1}{x_i} = \frac{1}{x_1}\cdot\frac{1}{x_2}\cdot\frac{1}{x_3} = \frac{1}{7}\cdot\frac{1}{1}\cdot\frac{1}{5} = \frac{1}{35} = 0.0286$$

7 $$\prod_{i=1}^{3} \frac{1}{x_i^2} = \frac{1}{x_1^2}\cdot\frac{1}{x_2^2}\cdot\frac{1}{x_3^2} = \frac{1}{[(7)(1)(5)]^2} = \frac{1}{(35)^2} = \frac{1}{1225} = 0.0008$$

8 $$\log\prod_{i=1}^{3} x_i = \sum_{i=1}^{3} \log x_i = \log x_1 + \log x_2 + \log x_3$$

$$= \log 7 + \log 1 + \log 5$$

$$= 0.8451 + 0 + 0.699 = 1.5441$$

9 $$\log\prod_{i=1}^{3} x_i^{a} = a\sum_{i=1}^{3} \log x_i = (6)\cdot(1.5441) = 9.2646$$

10 $$\log\prod_{i=1}^{3} x_i^{a\, y_i} = a\sum_{i=1}^{3} y_i \log x_i$$

$$= a\cdot\left[y_1 \log x_1 + y_2 \log x_2 + y_3 \log x_3\right]$$

$$= 6\cdot\left[2\cdot\log 7 + 0\cdot\log 1 + 4\cdot\log 5\right]$$

$$= 6\cdot\left[2\cdot 0.8451 + 0\cdot 0 + 4\cdot 0.699\right] = 6\cdot 4.4862 = 26.9172$$

11 $$\prod_{i=1}^{3}\sqrt{x_i} = \sqrt{x_1}\cdot\sqrt{x_2}\cdot\sqrt{x_3} = \sqrt{7}\cdot\sqrt{1}\cdot\sqrt{5} = (2.6457)(1)(2.2361) = 5.9161$$

$$\textit{or} \qquad = \sqrt{(7)(1)(5)} = \sqrt{35} = 5.9161$$

12 $$\prod_{i=1}^{3} x_i^a = x_1^a \cdot x_2^a \cdot x_3^a = 7^6 \cdot 1^6 \cdot 5^6$$

$$= (117649)(1)(15625) = 1838265625$$

or $$= (x_1 \cdot x_2 \cdot x_3)^a = (7 \cdot 1 \cdot 5)^6 = (35)^6 = 1838265625$$

13 $$\prod_{i=1}^{3} x_i y_i^2 = (x_1 y_1^2)(x_2 y_2^2)(x_3 y_3^2)$$

$$= (7 \cdot 2^2)(1 \cdot 0^2)(5 \cdot 4^2) = (28)(0)(80) = 0$$

14 $$\prod_{i=1}^{3} (x_i - y_i)^2 = (x_1 - y_1)^2 (x_2 - y_2)^2 (x_3 - y_3)^2$$

$$= (7-2)^2 (1-0)^2 (5-4)^2 = 5^2 \cdot 1^2 \cdot 1^2 = 25$$

15 $$\prod_{i=1}^{3} (x_i - a)^2 = (x_1 - a)^2 (x_2 - a)^2 (x_3 - a)^2$$

$$= (7-6)^2 (1-6)^2 (5-6)^2 = (1)^2 (-5)^2 (-1)^2 = 1 \cdot 25 \cdot 1 = 25$$

تمارين الفصل الثالث

1: إذا توفرت لديك البيانات الآتية:

x_i	-3	0	3	6	9

وأن k = 5 جد ناتج ما يأتي:

7	6	5	4	3	2	1
$\frac{1}{\sum_{i=1}^{n} x_i^2}$	$\frac{1}{\sum_{i=1}^{n} x_i}$	$\sum_{i=1}^{n}\left(\frac{1}{x_i}\right)^2$	$\sum_{i=1}^{n}\left(\frac{1}{x_i}\right)$	$\left(\sum_{i=1}^{n} x_i\right)^2$	$\sum_{i=1}^{n} x_i^2$	$\sum_{i=1}^{n} x_i$
14	13	12	11	10	9	8
$\frac{k}{\sum_{i=1}^{n} x_i}$	$\sum_{i=1}^{n} kx_i$	$\sum_{i=1}^{n} k$	$\sqrt{\sum_{i=1}^{n} x_i}$	$\sum_{i=1}^{n} \sqrt{x_i}$	$\log \sum_{i=1}^{n} x_i$	$\sum_{i=1}^{n} \log x_i$

2: أفرض لديك سلسلتين من البيانات الآتية:

x_i	0	2	4	6	8	10
y_i	1	3	5	7	9	11

s = 3 ، z = 5 ، جد ناتج مايلي:

5	4	3	2	1
$\prod_{i=1}^{z}(x_i - s)^2$	$\prod_{i=1}^{s}(x_i - y_i)^2$	$\prod_{i=1}^{n} x_i y_i^2$	$\prod_{i=1}^{n} x_i^z$	$\prod_{i=1}^{n} \sqrt[s]{x_i y_i}$

1-3 : إذا كانت d_i تمثل الإنفاق الشهري للأسرة في حين يمثل الدخل g_i لعينـة عـشوائية كانـت كما يلي:

d_i	105	206	300	121	238	100
g_i	200	300	500	340	340	110

وأن a = 3 جد مايلي:

7	6	5	4	3	2	1
$\frac{a\,n}{\sum_{i=1}^{n} g_i}$	$\sum_{i=1}^{a} n d_i$	$\sum_{i=1}^{n} a$	$\sqrt{\sum_{i=1}^{n} g_i}$	$\sum_{i=1}^{a}\sqrt{d_i}$	$\log\sum_{i=1}^{n} g_i$	$\sum_{i=1}^{n}\log d_i$
13	12	11	10		9	8
$\sum_{i=1}^{n}(d_i - g_i)^2$	$\sum_{i=1}^{n}(g_i - d_i)$	$\sum_{i=1}^{a}(g_i + a)$	$\sum_{i=1}(g_i + a)(d_i + n)$		$\sum_{i=1}^{n} d_i g_i^2$	$\sum_{i=1}^{n} d_i g_i$

3: أفرض لديك سلسلتين من البيانات الآتية:

x_i	3	5	8	2
y_i	2	6	4	5

والمطلوب إيجاد قيمة ما يلي:

5	4	3	2	1
$\log \prod_{i=1}^{2} x_i^{4y_i}$	$\log \prod_{i=1}^{n} x_i^3$	$\log \prod_{i=1}^{2} x_i$	$\prod_{i=1}^{n} \frac{1}{x_i^2}$	$\prod_{i=1}^{3} \frac{1}{x_i}$
10	9	8	7	6
$\prod_{i=1}^{n} (x_i - 12)^2$	$\prod_{i=1}^{n} (x_i - y_i)^2$	$\prod_{i=1}^{n} x_i y_i^2$	$\prod_{i=1}^{n} x_i^5$	$\prod_{i=1}^{n} \sqrt{x_i}$



الفصل الرابع

مقاييس النزعة المركزية

Measures of Central Tendency

الفصل الرابع

مقاييس النزعة المركزية

Measures of Central Tendency

1.4: مقدمة:

كثيراً ما نحتاج إلى حساب بعض المؤشرات التي تلخص البيانات بأقل قدر ممكن من التفاصيل أو نموذج يمثل المجموعة الإحصائية ومفرداتها أو معيار تقاس بالنسبة أليه هذه المفردات وتقارن بواسطته المجموعة كلها بالنسبة إلى المجموعات الإحصائية الأخرى. على هذا الأساس يمكن تمثيل مجموعة من البيانات بقيمة واحدة فقط الهدف من ذلك إعطاء صورة سريعة عن ماهية تلك المجموعة من خلال إيجاد عدد يمثل قيمها، أن المقياس الذي يختص بتحديد هذا العدد يسمى مقياس النزعة المركزية أو مقياس توسط (متوسط) هذا العدد يميل لأن يقع في وسط تلك المجموعة من البيانات في حال ترتيبها حسب صغرها أو كبرها، أي أن هذا العدد يؤول لإن يتمركز وسط المجموعة التي أحتسبت منها ولذلك سميت (مقاييس النزعة المركزية) والتي تفيد في دراسة خصائص المجتمع من خلال خصائص العينة التي تعتبر المتوسطات واحدة منها. وسنتناول في هذا الفصل أهم مقاييس النزعة المركزية للبيانات المبوبة وغير المبوبة.

2.4: الوسط الحسابي Mean:

يعتبر أهم مقاييس النزعة المركزية على الإطلاق لما يمتاز به من خصائص جيدة وسهولة في الحساب جعلته يقف في مقدمة مقاييس النزعة المركزية ويعد الوسط الحسابي المقياس الأوسع إستخداماً لأنه يأخذ جميع القيم بنظر الإعتبار عند حسابه ولايحتاج إلى ترتيب البيانات، ولكن من عيوبه أنه يتأثر بالقيم الشاذه أو المتطرفة للبيانات ويمكن حسابه للبيانات غير المبوبة والبيانات المبوبة وكما يلي:

أ- الوسط الحسابي للبيانات غير المبوبة:

يحسب الوسط الحسابي للبيانات غير المبوبة من خلال جمع كل قيم مشاهدات (القياسات) العينة $(x_1, x_2, \ldots, x_n)$ المأخوذه من المجتمع لظاهرة ما وقسمتها على حجم هذه العينة n، لذلك فإن الوسط الحسابي $\overline{x}$ لهذه العينة هو كما يلي:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \cdots \qquad (4.1)$$

إن الوسط الحسابي $\overline{x}$ هو تقدير غير متحيز لمعدل مشاهدات المجتمع الذي أختيرت منه هذه العينة، وأن معدل المجتمع يرمز له عادة μ ويمكن حسابه من خلال الصيغة الآتية:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} \qquad \cdots \qquad (4.2)$$

مثال (1.4)**: أفرض لديك البيانات التالية التي تمثل درجات عينة من الطلبة عددها 15 طالب في مادة الإقتصاد، يطلب إيجاد متوسط درجة الطالب في هذه العينة:**

x_i	65	66	58	56	75	43	25
56	90	78	32	59	35	42	80

الحل:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1}{15}(65 + 66 + \cdots + 80) = \frac{860}{15} = 57.3333$$

هذا يعني أن متوسط درجة الطلبة في مادة الإقتصاد هو 57.3333 وإذا رتبنا البيانات أعلاه تصاعدياً فإن هذه القيمة $\overline{x}$ ستقع في منتصف هذه البيانات تقريباً.

مثال (2.4)**: البيانات التالية تمثل عدد أفراد عينة من الأسر عـددها 10 أسر، يطلـب إيجـاد متوسـط عـدد أفراد الأسرة.**

x_i	3	7	8	12	8	2	5	6	5	4

الحل:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1}{10}(3 + 7 + \cdots + 4) = \frac{60}{10} = 6$$

وهذا يعني أن متوسط عدد أفراد الأسرة هو ستة أفراد.

مثال (3.4)**: البيانات التالية تمثل الدخل الشهري (بألآف الدنانير) لعينة تتكون من (6) عوائل مأخوذه من أربيل وهي كما يلي:**

y_i	300	800	1200	790	200	500

والمطلوب إيجاد الوسط الحسابي للدخل الشهري لهذه العينة.

الحل:

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{1}{6}(300 + 800 + \cdots + 500) = \frac{3790}{6} = 631.6667$$

وهذا يعني أن متوسط الدخل الشهري للأسرة لهذه العينة هو 631.6667 دينار.

ب- الوسط الحسابي للبيانات المبوبة:

يتم إيجاد متوسط البيانات المبوبة من خلال حساب مراكز الفئـات x_i وضرب كـلاً منهـا بمـا يقابلها من التكرارات f_i ومن ثم جمع نتائج حاصل الضرب وقسمتها على مجموع التكرارات، أي أن:

$$\overline{x} = \frac{\sum_{i=1}^{m} f_i x_i}{\sum_{i=1}^{m} f_i} \qquad \cdots \quad (4.3)$$

مثال (4.4)**: الجدول التالي يبين كميات الإنتاج (بالقطعة) لخمسين عاملاً في مصنع أربيل وعدد العـمال في كل فئة من الفئات.**

فئات الإنتاج	100-	110-	120-	130-	140-	150-	160-	170- 180
عدد العمال	4	7	8	10	7	6	5	3

والمطلوب إيجاد متوسط إنتاجية العامل من القطع.

الحل: نقوم أولاً بتكوين جدول يبين فئات الإنتاج مقابل عـدد العـمال f_i وحـساب مراكـز الفئـات مـن خلال ما يلي:

بالنسبة إلى مركز الفئة الأولى هو:

$$x_1 = \frac{L.L + U.L}{2} = \frac{100 + 110}{2} = 105$$

بالنسبة إلى مركز الفئة الثانية هو:

$$x_2 = \frac{L.L + U.L}{2} = \frac{110 + 120}{2} = 115$$

وهكذا البقية...، ويتم من خلال هذا الجدول أيضاً حساب حاصل ضرب مراكز الفئات في ما يقابلها من تكرارات والجدول الآتي يوضح ذلك:

فئات الإنتاج	التكرار f_i (عدد العمال)	مراكز الفئات x_i	$f_i x_i$
-100	4	105	420
-110	7	115	805
-120	8	125	1000
-130	10	135	1350
-140	7	145	1015
-150	6	155	930
-160	5	165	825
180-170	3	175	525
المجموع	50		6870

لذلك فإن:

$$\overline{x} = \frac{\sum_{i=1}^{m} f_i x_i}{\sum_{i=1}^{m} f_i} = \frac{\sum_{i=1}^{8} f_i x_i}{\sum_{i=1}^{8} f_i} = \frac{f_1 x_1 + f_2 x_2 + \cdots f_8 x_8}{f_1 + f_2 + \cdots + f_8}$$

$$= \frac{(4)\cdot(105) + (7)\cdot(115) + \cdots + (3)\cdot(175)}{4 + 7 + \cdots + 3} = \frac{6870}{50} = 137.4$$

وهذا يعني أن متوسط إنتاج العمال هو 137.4 قطعة.

مثـــال (5.4)**: بـــالإعتماد عـــلى جـــدول المثـــال** (1.2) **لدراســـة أجـــور العـــاملين اليوميـــة (بـــآلاف**

الدنانير) للمواد الإنشائية وكانت كما يلي:

تسلسل الفئة	الفئات (الأجور)	التكرار f_i (عدد العمال)	مراكز الفئات x_i
1	29-26	5	27.5
2	33-30	12	31.5
3	37-34	11	35.5
4	41-38	9	39.5
5	45-42	5	43.5
6	49-46	3	47.5

والمطلوب إيجاد متوسط الأجر اليومي للعاملين في هذا المصنع؟

الحل: من خلال الجدول التكراري يمكن حساب المتوسط وكما يلي:

$$\bar{x} = \frac{\sum_{i=1}^{m} f_i x_i}{\sum_{i=1}^{m} f_i} = \frac{\sum_{i=1}^{6} f_i x_i}{\sum_{i=1}^{6} f_i} = \frac{f_1 x_1 + f_2 x_2 + \cdots + f_6 x_6}{f_1 + f_2 + \cdots + f_6}$$

$$= \frac{(5)\cdot(27.5)+(12)\cdot(31.5)+\cdots+(3)\cdot(47.5)}{5+12+\cdots+3} = \frac{1621.5}{45} = 36.0333$$

متوسط الأجر اليومي للعاملين في هذا المصنع هو 36.0333 ألف دينار.

مثال (6.4)**:** بالإعتماد على الجدول التكراري للمثال (2.2) للبيانات التي تمثل كميات إنتاج الحنطة في الدونم الواحد إلى (60) قطعة زراعية، والمطلوب حساب متوسط كميات إنتاج الحنطة في الدونم الواحد؟

الحل: من خلال الجدول التكراري للمثال (2.2) يمكن الحصول على ما يلي:

الفئات (كمية إنتاج الحنطة بالطن)	التكرار f_i (عدد قطع الأراضي)	مراكز الفئات x_i	$f_i x_i$
26.70-17.83	1	22.27	22.27
35.56-26.70	14	31.13	435.82
44.42-35.56	20	39.00	780
53.28-44.42	8	48.86	390.88
62.15-53.28	7	57.73	404.11
71.02-62.15	6	66.59	399.54
79.88-71.02	4	75.46	301.84
المجموع	60		2734.46

من خلال الجدول التكراري يمكن حساب المتوسط وكما يلي:

$$\bar{x} = \frac{\sum_{i=1}^{m} f_i x_i}{\sum_{i=1}^{m} f_i} = \frac{\sum_{i=1}^{7} f_i x_i}{\sum_{i=1}^{7} f_i} = \frac{f_1 x_1 + f_2 x_2 + \cdots + f_7 x_7}{f_1 + f_2 + \cdots + f_7}$$

$$= \frac{(1)\cdot(22.27) + (14)\cdot(31.13) + \cdots + (4)\cdot(75.46)}{1 + 14 + \cdots + 4} = \frac{2734.46}{60} = 45.574$$

وهذا يعني أن متوسط كميات إنتاج الحنطة في الدونم الواحد هو 45.5743 طن.

ملاحظة: الطريقة أعلاه يمكن استخدامها في حالة تساوي أطوال الفئات وحالة عدم

تساويها وهنالك أنواع أخرى لامجال لذكرها هنا.

مميزات وعيوب الوسط الحسابي:

فيما يلي أهم مميزات وعيوب الوسط الحسابي:

أ- **مميزات الوسط الحسابي:** يتميز مقياس الوسط الحسابي بعدة أمور منها:

بساطة فكرته وسهولة حسابه.

1- يأخذ جميع البيانات بنظر الإعتبار عند حسابه.

2- يمكن حسابه بالأعتماد على العمليات الجبرية.

ب- عيوب الوسط الحسابي: يعاب على الوسط الحسابي ما يلي:

1. لايمكن تحديده بالنظر إلى البيانات أو هندسياً.
2. لايمكن حسابه للبيانات الوصفية غير القابلة للترتيب (الأسمية-Nominal) مثل الجنس، القومية،..الخ.
3. لايمكن حسابة عند وجود قيم مفقودة إلا بعد تقديرها.
4. يتأثر كثيراً بالقيم الشاذة أو المتطرفة.

خصائص الوسط الحسابي:

1- إن مجموع إنحرافات قيم المتغير x عن وسطها الحسابي الذي أحتسب منها يكون مساوياً للصفر، أي أن:

2-

$$d_i = x_i - \bar{x}$$

$$\Rightarrow \sum_{i=1}^{n} d_i = \sum_{i=1}^{n} \left(x_i - \bar{x}\right) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} = \sum_{i=1}^{n} x_i - n\,\bar{x}$$

$$= n\bar{x} - n\bar{x} = 0$$

2- إن مجموع مربعات إنحرافات قيم x عن وسطها الحسابي الذي أحتسب منها يكون أقل ما يمكن، أي أن:

$$\sum_{i=1}^{n} d_i^2 < \varepsilon$$

3.4: الوسط الحسابي الموزون (المرجح) Weighted Mean:

هنالك الكثير من الحالات تكون بعض المفردات أكثر أهمية من الأخرى، مما يستوجب ذلك أخذه بنظر الإعتبار لدى حساب الوسط الحسابي فمثلاً عند حساب معدل درجات الطالب المتخرج من الكلية فإن الأمر يتطلب الأخذ بنظر الإعتبار عدد الساعات الأسبوعية المخصصة لكل مادة من المواد الدراسية التي تدخل في حساب المعدل، وهذا يعني أن هنالك عملية ترجيح لقياسات المفردات بأوزان Weights تمثل أهمية كل منها، هذه الأوزان غالباً ماتكون محددة مسبقاً. ويمكن حسابه للبيانات غير المبوبة والبيانات المبوبة وكما يلي:

أ- الوسط الحسابي الموزون للبيانات غير المبوبة:

أفرض أن $\left(x_1, x_2, \ldots, x_n\right)$ تمثل قياسات عينة من المفردات عددها n لبيانات غير مبوبة وأن $\left(w_1, w_2, \ldots, w_n\right)$ تمثل أوزان هذه المفردات، عندئذ يعرف الوسط الحسابي الموزون (المرجح) على النحو الآتي:

$$\overline{x}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \qquad \cdots(4.4)$$

ويلاحظ مما تقدم أن الوسط الحسابي المرجح هو حالة أكثر عمومية من الوسط الحسابي الإعتيادي، أي مانعنيه أن الوسط الحسابي الإعتيادي هو حالة خاصة من الوسط الحسابي المرجح عندما ينظر الى كافة المفردات بنفس الأهمية (الوزن) فعلى إفتراض أن:

$$w_i = w \quad i.e \quad w_1 = w_2 = \ldots = w_n = w \quad \forall \quad i$$

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} = \frac{\sum_{i=1}^{n} w x_i}{\sum_{i=1}^{n} w} = \frac{w \sum_{i=1}^{n} x_i}{w \sum_{i=1}^{n} 1} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$$

يلاحظ أن الناتج الأخير ماهو الى الوسط الحسابي الإعتيادي.

مثال (7.4)**: كانت درجات أحد طلبة قسم الإقتصاد في المواد المقررة حسب الساعات الأسبوعية المحددة لكل مادة هي كما يلي:**

70	87	86	84	88	75	80	52
3	3	3	3	3	2	2	2

والمطلوب إيجاد الوسط الحسابي الموزون؟

الحل: نلاحظ أن هنالك إختلاف في عدد الساعات الأسبوعية المحددة لكل مادة لذلك يستوجب حساب المتوسط الموزون (المرجح) وكما يلي:

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} = \frac{(3 \times 70) + (3 \times 87) + \ldots + (2 \times 52)}{3 + 3 + \ldots + 2} = \frac{1659}{21} = 79$$

وهذا يعني أن معدل الطالب هو 79 في حين إذا لم يؤخذ عدد الساعات بنظر الإعتبار (وهذا غير صحيح طبعاً) فإن معدل درجات الطالب وفق الصيغة الإعتيادية هو 77.75 ونلاحظ أنه يفرق بشكل كبير عن المعدل الموزون.

مثال (8.4)**: فيما يلي الراتب الشهري (بآلآف الدنانير) لجميع الموظفين في جامعة صلاح**

الدين/ أربيل:

الدخل الشهري	عدد الموظفين	نوع العمل
2340	95	إداري
1800	180	أكاديمي
850	253	فني
350	67	خدمي

والمطلوب حساب معدل الدخل الشهري لموظفي جامعة صلاح الدين؟

الحل: يمكن حساب معدل الدخل الشهري لموظفي جامعة صلاح الدين وكما يلي:

$$\overline{x}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} = \frac{(95 \times 2340) + (180 \times 1800) + (253 \times 850) + (67 \times 350)}{95 + 180 + 253 + 67}$$

$$= \frac{784800}{595} = 1318.9916$$

ب- الوسط الحسابي الموزون للبيانات المبوبة:

يمكن حساب المتوسط الموزون للبيانات المبوبة من خلال الصيغة الآتية:

$$\overline{x}_w = \frac{\sum_{i=1}^{m} w_i f_i x_i}{\sum_{i=1}^{m} w_i f_i} \qquad \cdots (4.5)$$

مثال (9.4)**: فيما يلي توزيع تكراري للإنتاج اليومي لمصنع أربيل من السكر موزع حسب عدد المكائن وعدد ساعات العمل المحددة حسب مواصفات المنشأ لتلك الماكنة، وكما يلي:**

فئات الإنتاج (كميات الإنتاج بالطن)	عدد المكائن	ساعات العمل المقررة
4-	8	12
8-	10	10
12-	12	12
16-	6	8
20-24	4	8

والمطلوب حساب متوسط إنتاجية الماكنة الواحدة في هذا المصنع.

الحل: لحساب متوسط إنتاجية الماكنة الواحدة نحتاج إلى تكوين الجدول الآتي:

الفئات	عدد المكائن (التكرار f_i)	ساعات العمل المقررة (الأوزان w_i)	مراكز الفئات x_i	$w_i f_i$	$w_i f_i x_i$
4-	8	12	6	96	576
8-	10	10	10	100	1000
12-	12	12	14	144	2016
16-	6	8	18	48	864
20-24	4	8	22	32	704
المجموع	40			420	5160

لذلك فإن متوسط إنتاجية الماكنة الواحدة هو كما يلي:

$$\bar{x}_w = \frac{\sum_{i=1}^{m} w_i f_i x_i}{\sum_{i=1}^{m} w_i f_i} = \frac{\sum_{i=1}^{5} w_i f_i x_i}{\sum_{i=1}^{m} w_i f_i} = \frac{5160}{420} = 12.2857$$ طن

ملاحظة: إن مزايا وعيوب وخصائص الوسط الحسابي الموزون هي مشابه تماماً للوسط الحسابي الإعتيادي.

4.4: الوسط التوافقي Harmonic Mean:

يستخدم الوسط التوافقي في حساب معدل سعر صرف العملة مقابل العملات الأجنبية الأخرى، أو في إيجاد المتوسط للمعدلات الزمنية المختلفة مثل (عدد الوحدات المنتجة من مصنع معين خلال فترة محددة) أو (إيجاد متوسط القراءة لمجموعة من الأفراد بدلالة عدد الكلمات في الدقيقة)،... الخ. ويمكن حسابه في حالة البيانات غير المبوبة وحالة البيانات المبوبة وكما يلي:

أ- الوسط التوافقي للبيانات غير المبوبة:

وهو عبارة عن مقلوب الوسط الحسابي لمقلوبات القيم والصيغة العامة له كما يلي:

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \qquad \cdots \qquad (4.6)$$

مع ملاحظة أن قيم المشاهدات x_i يجب أن لاتساوي صفر.

مثال (10.4): **جد الوسط التوافقي للبيانات الآتية:**

x	4	5	3	4	7	8	5	10

الحل:

نجد أولاً مقلوب القيم وكما يلي:

$$\frac{1}{x_i} = 0.25 \quad 0.2 \quad 0.33 \quad 0.25 \quad 0.14 \quad 0.13 \quad 0.2 \quad 0.1$$

فإن الوسط التوافقي هو كما يلي:

$$\therefore H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{8}{\sum_{i=1}^{8} \frac{1}{x_i}} = \frac{8}{0.25 + 0.2 + \ldots + 0.1} = \frac{8}{1.6} = 5$$

ب- الوسط التوافقي للبيانات المبوبة:

يمكن حساب الوسط التوافقي للبيانات المبوبة ذات الفئات المتساوية وحتى غير المتـساوية مـن خلال الصيغة الآتية:

$$H = \frac{\sum_{i=1}^{m} f_i}{\sum_{i=1}^{m} \frac{f_i}{x_i}} \quad \cdots \quad (4.7)$$

مثال (11.4)**: جد الوسط التوافقي للتوزيع التكراري الآتي:**

الفئات	التكرار
50-	13
60-	18
70-	25
80-	12
90-	7
100-110	5

لحساب الوسط التوافقي يتم تكوين الجدول الآتي:

الفئات	مراكز الفئات x_i	التكرار f_i	$\frac{f_i}{x_i}$
50-	55	13	0.2364
60-	65	18	0.2769
70-	75	25	0.3333
80-	85	12	0.1412
90-	95	7	0.0737
100-110	105	5	0.0476
المجموع		80	1.1091

من خلال الجدول أعلاه يمكن حساب الوسط التوافقي وكما يلي:

$$H = \frac{\sum_{i=1}^{m} f_i}{\sum_{i=1}^{m} \frac{f_i}{x_i}} = \frac{\sum_{i=1}^{6} f_i}{\sum_{i=1}^{6} \frac{f_i}{x_i}} = \frac{80}{1.1091} = 72.1306$$

مميزات وعيوب الوسط التوافقي:

فيما يلي أهم مميزات وعيوب الوسط التوافقي:

أ- **مميزات الوسط التوافقي:** يتميز مقياس الوسط التوافقي بعدة أمور منها:

1- بساطة فكرته.

2- يأخذ جميع البيانات بنظر الإعتبار عند حسابه.

3- يمكن حسابه بالأعتماد على العمليات الجبرية.

ب- عيوب الوسط التوافقي: يعاب على الوسط التوافقي ما يلي:

1- هنالك بعض الصعوبة في حسابه.

2- لايمكن حسابه للبيانات الوصفية غير القابلة للترتيب.

3- لايمكن حسابه عند وجود قيم مفقودة إلا بعد تقديرها.

4- يتأثر كثيراً بالقيم الشاذة أو المتطرفة.

5- لايمكن حسابه عندما تكون أحد قيم البيانات مساوٍ للصفر أو مركز الفئة مساوٍ للصفر.

5.4: الوسط التربيعي Quadratic Mean:

يمكن إيجاده في حالة البيانات غير المبوبة وحالة البيانات المبوبة وكما يلي:

أ- الوسط التربيعي للبيانات غير المبوبة:

الوسط التربيعي للبيانات غير المبوبة هو الجذر التربيعي الموجب لمتوسط مربعات قيم المشاهدات، أي أن:

$$Q = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}} \qquad \cdots \qquad (4.8)$$

مثال (12.4)**: جد الوسط التربيعي لدرجات حرارة مدينة أربيل أثناء فصل الشتاء لسبعة أيام متتالية وكما يلي:**

1	0	-1	-2	3	5	4	x_i

الحل:

$$Q = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}} = \sqrt{\frac{\sum_{i=1}^{7} x_i^2}{7}} = \sqrt{\frac{4^2 + 5^2 + \cdots + 1^2}{7}}$$

$$= \sqrt{\frac{56}{7}} = \sqrt{8} = 2.8284$$

أي أن الوسط التربيعي لدرجات حرارة مدينة أربيل لتلك الفترة هي 2.8284 درجة مئوية.

ب- الوسط التربيعي للبيانات المبوبة:

يمكن الحصول على الوسط التربيعي للبيانات المبوبة من خلال الصيغة الآتية:

$$Q = \sqrt{\frac{\sum_{i=1}^{m} f_i x_i^2}{\sum_{i=1}^{m} f_i}} \quad \cdots \quad (4.9)$$

مثال (13.4)**: جد الوسط التربيعي للجدول التكراري الآتي:**

الفئات	التكرار f_i
10-	5
20-	6
30-	8
40-	7
50-60	4

الحل: لإيجاد الوسط التربيعي يتم تكوين الجدول الآتي:

الفئات	مراكز الفئات x_i	x_i^2	التكرار f_i	$f_i x_i^2$
10-	15	225	5	1125
20-	25	625	6	3750
30-	35	1225	8	9800
40-	45	2025	7	14175
50-60	55	3025	4	12100
المجموع			30	40950

لذلك فإن الوسط التربيعي هو كما يلي:

$$Q = \sqrt{\frac{\sum_{i=1}^{m} f_i x_i^2}{\sum_{i=1}^{m} f_i}} = \sqrt{\frac{\sum_{i=1}^{5} f_i x_i^2}{\sum_{i=1}^{5} f_i}}$$

$$= \sqrt{\frac{40950}{30}} = \sqrt{1365} = 36.9459$$

ملاحظة: مميزاته وعيوبه مشابه تماماً للوسط التوافقي ماعدا إمكانية حسابه عندما تكون قيمة المشاهدة مساوية للصفر.

6.4: الوسط الهندسي Geometric Mean:

الوسط الهندسي مهم جداً في الدراسات السكانية وخصوصاً عند حساب معدلات نمو السكان وكذلك في تكوين الأرقام القياسية (Index numbers) ويمكن حسابه في حالة

البيانات غير المبوبة وحالة البيانات المبوبة وكما يلي:

أ- الوسط الهندسي للبيانات غير المبوبة:

يعرف على أنه الجذر الموجب ذو الرتبة n لحاصل ضرب قياسات مجموعة من المشاهدات ببعضها، أي أن:

$$G = \sqrt[n]{\prod_{i=1}^{n} x_i} = \left(\prod_{i=1}^{n} x_i \right)^{\frac{1}{n}} \quad \cdots \quad (4.10)$$

ويمكن إيجاده بطريقة أخرى من خلال أخذ اللوغارتم للأساس 10 وكما يلي:

$$Log_{10} G = Log_{10} \left(\prod_{i=1}^{n} x_i \right)^{\frac{1}{n}}$$

$$= \frac{1}{n} \sum_{i=1}^{n} Log_{10} x_i$$

$$\therefore \quad G = anti - Log_{10} \left(\frac{1}{n} \sum_{i=1}^{n} Log_{10} x_i \right) \quad \cdots \quad (4.11)$$

مثال (14.4)**: جد الوسط الهندسي للبيانات الآتية:**

x_i	5	10	15	20	25

الحل:

$$G = \sqrt[n]{\prod_{i=1}^{n} x_i} = \left(\prod_{i=1}^{n} x_i \right)^{\frac{1}{n}}$$

$$\therefore G = \sqrt[5]{5 \times 10 \times \ldots \times 25} = \sqrt[5]{375000}$$

$$= (375000)^{\frac{1}{5}} = 13.0259$$

أو الحل بطريقة أخرى:

$$G = anti - Log_{10}\left(\frac{1}{n}\sum_{i=1}^{n} Log_{10}x_i\right)$$

$$= anti - Log_{10}\left(\frac{1}{5}\left(Log_{10}(5) + Log_{10}(10) + \ldots + Log_{10}(25)\right)\right)$$

$$= anti - Log_{10}\left(\frac{1}{5}\left(0.699 + 1 + \ldots + 1.3979\right)\right)$$

$$= anti - Log_{10}\left(\frac{1}{5}\left(5.574\right)\right)$$

$$= anti - Log_{10}\left(1.1148\right) = 13.0259$$

ب- الوسط الهندسي للبيانات المبوبة:

يمكن الحصول على الوسط الهندسي للبيانات المبوبة من خلال الصيغة الآتية:

$$G = \sqrt[\sum_{i=1}^{m} f_i]{\prod_{i=1}^{m} x_i^{f_i}} = \left(\prod_{i=1}^{m} x_i^{f_i}\right)^{\frac{1}{n}} ; \quad n = \sum_{i=1}^{m} f_i \quad \cdots \quad (4.12)$$

ويمكن إيجاده بطريقة أخرى من خلال أخذ اللوغارتم للأساس 10 وكما يلي:

$$\log_{10} G = \frac{1}{n}\log_{10}\left(\prod_{i=1}^{m} x_i^{f_i}\right) = \frac{1}{n}\sum_{i=1}^{m} f_i \log_{10} x_i$$

$$\therefore \quad G = anti - \log_{10}\left(\frac{1}{n}\sum_{i=1}^{m} f_i \log_{10} x_i\right) \quad \cdots \quad (4.13)$$

مثال (15.4)**: جد الوسط الهندسي للجدول التكراري الآتي:**

التكرار f_i	الفئات
2	0-
4	10-
5	20-
6	30-
3	40-50

الحل: لإيجاد الوسط الهندسي للبيانات المبوبة يتم تكوين الجدول الآتي:

$f_i \log_{10} x_i$	$\log_{10} x_i$	التكرار f_i	مراكز الفئات x_i	الفئات
1.3980	0.699	2	5	0-
4.7044	1.1761	4	15	10-
6.9895	1.3979	5	25	20-
9.2646	1.5441	6	35	30-
4.9596	1.6532	3	45	40-50
27.3161		20		المجموع

لذلك فإن الوسط الهندسي هو كما يلي:

$$G = anti-\log_{10}\left(\frac{1}{n}\sum_{i=1}^{m} f_i \log_{10} x_i\right)$$
$$= anti-\log_{10}\left(\frac{1}{20}\sum_{i=1}^{5} f_i \log_{10} x_i\right)$$
$$= anti-\log_{10}\left(\frac{27.3161}{20}\right) = anti-\log_{10}(1.3658)$$
$$\therefore \quad G = 23.2167$$

ملاحظة: يمكن إيجاد الوسط الهندسي عندما تكون أطوال الفئات متساوية وحتى إذا كانت غير متساوية، وأن مميزاته وعيوبه مشابه تماماً للوسط التوافقي ماعدا كونه لايمكن حساب عندما تكون أحد قيم المشاهدات ذات قيمة سالبة.

7.4: المنوال The Mode:

يمثل المنوال القيمة التي تتكرر أكثر من غيرها من المشاهدات أو الصفة الأكثر شيوعاً لذا فإنه يفضل عندما يكون المطلوب معرفة ذوق المستهلك (أو مواصفات) حول سلعة معينة، ويمكن إيجاده في حالة البيانات غير المبوبة وحالة البيانات المبوبة وكما يلي:

أ- المنوال للبيانات غير المبوبة:

المنوال هو القيمة الأكثر تكراراً في مجموعة بيانات عينة، وإذا كانت كل القيم مكررة بنفس العدد من التكرارات فإنه في هذه الحالة لايوجد منوال. في حين يمكن أن يكون هنالك منوال واحد أو أكثر من منوال لمجموعة بيانات عينة إذا تكررت قيمة واحدة أو أكثر بنفس عدد التكرارات.

مثال (16.4)**: معرض شيشار للسيارات كانت مبيعاته من سيارة فورد في الفترة الأخيرة حسب الألوان كما يلي:**

حمراء	بيضاء	سوداء	بيضاء	صفراء
بيضاء	سوداء	حمراء	بيضاء	خضراء

والمطلوب معرفة المنوال وماذا تنصح هذه الشركة؟

الحل: من خلال المبيعات نلاحظ أن السيارة ذات اللون الأبيض كانت أكثر تكراراً (تكررت أربعة مرات) من بقية الألوان لذلك فإن المنوال هو السيارة ذات اللون الأبيض وننصح الشركة بالإكثار من إستيراد اللون الأبيض لأنه الأكثر مبيعاً.

مثال (17.4)**: جد المنوال للبيانات الآتية:**

5	4	7	3	5	6	2	5	3	4	5	x_i

الحل: من خلال البيانات أعلاه نلاحظ أن المنوال هو 5 لإنه تكرر (تكرر 4 مرات) أكثر من غيره من القيم.

مثال (18.4)**: جد المنوال للبيانات الآتية:**

10	7	5	3	2	1	x_i

الحل: لايوجد منوال للبيانات أعلاه لأنه لايوجد عدد تكرر أكثر من غيره.

ب- المنوال للبيانات المبوبة:

هنالك عدة طرائق في حساب المنوال للبيانات المبوبة حسب نوعها وهنا سنوضح أبسط طريقة وهي أن المنوال هو تلك القيمة التي تمثل مركز الفئة المقابلة لأكبر تكرار في جدول التوزيع التكراري، في حين هنالك عدة طرائق لامجال لذكرها هنا مثل طريقة بيرسون أو طريقة العزوم أو...الخ.

مثال (19.4)**: جد المنوال للجدول التكراري الآتي:**

الفئات	التكرار f_i
0-	1
10-	3
20-	7
30-	5
40-50	2

الحل: من خلال الجدول أعلاه نلاحظ أن أكبر تكرار هو 7 يقابل الفئة (20-30) أي أن المنوال يساوي مركز هذه الفئة وهو 25

ملاحظة: يمتاز المنوال بسهولة حسابه ولايتأثر بالقيم الشاذة أو المتطرفة وهو أفضل مقياس للبيانات من النوع الوصفي، إلا أنه يعاب عليه أنه لا يأخذ جميع البيانات بنظر الإعتبار عند حسابه ولايخضع لقاعدة جبرية ثابته في كل حالات حسابه.

8.4: الوسيط The Median:

الوسيط هو القيمة التي يقع ترتيبها وسط المجموعة عند ترتيب هذه القيم ترتيباً تصاعدياً أو تنازلياً، ويمكن ايجاده في حالة البيانات المبوبة وفي حالة البيانات غير المبوبة وكما يلي:

أ- الوسيط للبيانات غير المبوبة:

في حالة البيانات غير المبوبة إذا كان عدد القيم n عدد فردي عندئذ قيمة الوسيط تمثل قيمة x بعد الترتيب والتي تسلسلها هو $\frac{n+1}{2}$ أما إذا كان عدد القيم n عدد زوجي عندئذ فإن قيمة الوسيط تمثل الوسط الحسابي لقيمتي x بعد الترتيب اللتين تسلسلهما على التوالي هو: $\frac{n}{2}$ و $\frac{n}{2}+1$

مثال (20.4)**: جد الوسيط لدرجات 9 طلاب في إمتحان معين وكما يأتي:**

x	55	62	53	70	68	65	63	79	80

الحل:

بمـــا أن عـــدد المـــشاهدات n هـــي عـــدد فـــردي فـــإن التسلـــسل المطلـــوب هـــو:

$\frac{n+1}{2} = \frac{9+1}{2} = 5$ لذلك فإن:

53	55	62	63	65	68	70	79	80	الترتيب التصاعدي
1	2	3	4	5	6	7	8	9	التسلسل

أي أن تسلسل القيمة الخامسة هي التي تمثل الوسيط أي الدرجة 65 = me

مثال (21.4)**: جد الوسيط لعمر الفرد في هذه المجموعة:**

x	20	22	19.5	26	24.5	27	28	29	18	20	23	25

الحل: ترتيب هذه القيم وفق ترتيب تصاعدي:

18	19.5	20	20	22	23	24.5	25	26	27	28	29	الترتيب التصاعدي
1	2	3	4	5	6	7	8	9	10	11	12	التسلسل

بما أن n هي عدد زوجي فإن التسلسل المطلوب هـو $\frac{n}{2}$ و $\frac{n}{2}+1$ أي $\frac{12}{2}$ و $\frac{12}{2}+1$ أي

6 و 7 وهما يحددان القيمتان 23 و 24.5 لذلك فإن الوسيط هو المعدل بينهما، أي أن:

$$me = \frac{23 + 24.5}{2} = 23.75 \quad \text{سنة}$$

ب- الوسيط للبيانات المبوبة:

يمكن إيجاد الوسيط للبيانات المبوية من خلال التوزيع التكراري الصاعد وإيجاد ترتيب الوسيط بحساب نصف التكرار الكلي وتحديد الفئة الوسيطية التي تقابل التكرار المتجمع الصاعد الذي يلي ترتيب الوسيط مباشرةً، واستخدام القانون الآتي في إيجاد الوسيط:

$$Me = L_i + \left[\frac{n/2 - F_i}{f_i} \right] \cdot w \qquad \cdots \quad (4.14)$$

حيث أن:

L_i تمثل الحد الأدنى لفئة الوسيط.

f_i تمثل تكرار فئة الوسيط.

F_i تمثل التكرار المتجمع الصاعد السابق لفئة الوسيط.

w طول فئة الوسيط ناقص واحد إذا كان المتغير من النوع المتقطع.

أما من جدول التوزيع التكرار النازل نستخدم القانون الآتي:

$$Me = L_i + \left[\frac{F_i' - n/2}{f_i} \right] \cdot w \qquad \cdots \quad (4.15)$$

حيث أن:

F_i' تمثل التكرار المتجمع النازل الذي يقابل الفئة الوسيطية.

مثال (22.4)**: جد الوسيط لجدول التوزيع التكراري في المثال** (4.2)**الذي حصلنا من خلاله على ما يلي:**

تسلسل الفئة	الحدود العليا للفئات	التكرار f_i (عدد العمال)	التكرار المتجمع الصاعد F_i
1	30	5	5
2	34	12	17
3	38	11	28
4	42	9	37
5	46	5	42
6	50	3	45

الحل: ترتيب الوسيط = 45 على 2 يساوي 22.5 وهو يقع بين17 و28 يقابل الفئة الوسيطة 34-38 أي أن:

$$Me = L_i + \left[\frac{n/2 - F_i}{f_i} \right] \cdot w$$

$$= 34 + \left[\frac{22.5 - 17}{11} \right] \cdot 3 = 35.5$$

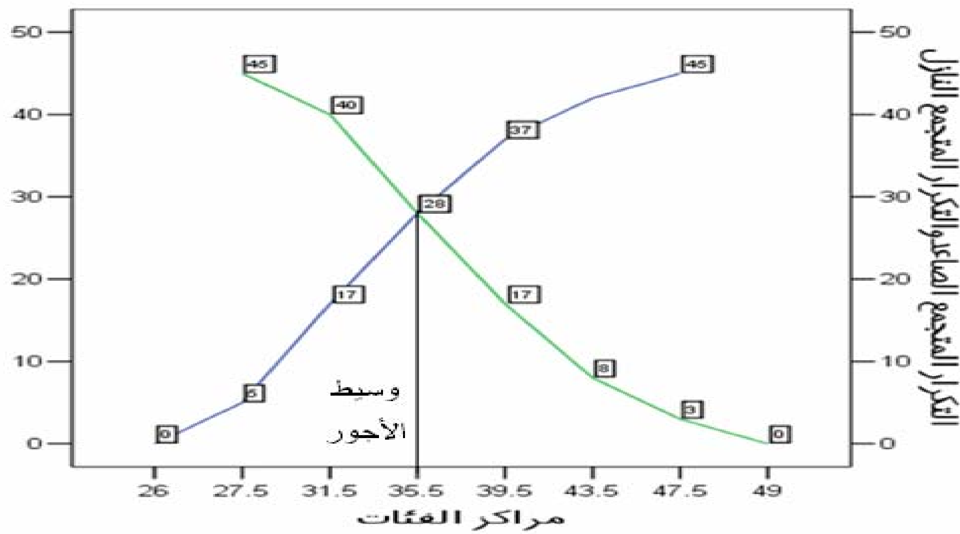
وباستخدام المتجمع التكراري النازل:

تسلسل الفئة	الحدود الدنيا للفئات	التكرار f_i (عدد العمال)	التكرار المتجمع النازل F'_i
1	26	5	45
2	30	12	40
3	34	11	28
4	38	9	17
5	42	5	8
6	46	3	3

$$Me = L_i + \left[\frac{n/2 - F_i'}{f_i} \right] \cdot w$$

$$= 34 + \left[\frac{28 - 22.5}{11} \right] \cdot 3 = 35.5$$

ويمكن تحديد قيمة الوسيط من خلال نقطة تقاطع التكرار الصاعد والنازل التي تقابل القيمة 35.5 وكما مبين في الشكل الآتي:



الشكل (1.4): المنحنى التكراري المتجمع الصاعد والنازل

إلى 45 عامل حسب أجورهم اليومية مع تحديد قيمة الوسيط

ملاحظة.1: يمتاز الوسيط بعدم تأثره بالقيم الشاذه وإمكانية إيجاده للقيم الوصفية القابلة للترتيب ولكن يعاب عليه أنه لايأخذ جميع القيم بنظر الإعتبار عند حسابه وهنالك بعض الصعوبه في حسابه.

ملاحظة.2: إذا كان التوزيع متماثل (يتبع التوزيع الطبيعي) فإن الوسط الحسابي والمنوال والوسيط تكون متساوية، وكلما يتباعد التوزيع عن التماثل تتباعد تلك القيم عن بعضها.

تمارين الفصل الرابع

4-1: ما المقصود بمقاييس النزعة المركزية؟

4-2: جد الوسط الحسابي للدخل الشهري لعينة مسحوبة من مجتمع اربيل وكما يلي:

520	740	950	650	500	200	900	1000	800	700	400

4-3: أحسب معدل درجات الحرارة لمدينة أربيل خلال سنة واحدة:

5	10	12	18	20	28	30	35	48	45	40	30

4-4: إذا كان الإستهلاك العائلي الشهري من مادة السكر في مدينة اربيل ودهوك كما يلي:

اربيل	7.8	7.50	6.5	5.9	2.70	9.06	10.3	13.9	12.4	15	12
دهوك	3.4	4.9	13	12	5.8	7.9	4.8	6.8	5.9	10	9.8

والمطلوب معرفة هل أن معدل الإستهلاك العائلي الشهري من مادة السكر في اربيل أقل من دهوك ؟

4-5: تمثل البيانات التالية المبالغ بالدينار التي تقاضاها 45 مندوباً من مندوبي المبيعات في أحد الأسابيع:

34	37	41	26	30	35	37	38	29
38	40	27	31	33	37	42	44	30
35	39	30	34	39	40	49	28	32
43	46	31	31	44	26	33	31	39
48	34	32	36	45	32	35	35	38

والمطلوب ما يلي:

أ- ضع هذه البيانات في جدول توزيع تكراري.

ب- أحسب الوسط الحسابي للبيانات المبوبة.

4-6: أحسب متوسط الأجور للجدول التكراري في المثال (1.2):

تسلسل الفئة	الفئات (الأجور)	التكرار f_i (عدد العمال)
1	29-26	5
2	33-30	12
3	37-34	11
4	41-38	9
5	45-42	5
6	49-46	3

4-7: مصنع في مدينة السليمانية مؤلف من ثلاثة أقسام كل منه يختص بإنتاج سلعة معينة وأن عدد العاملين ومتوسط الأجر الشهري (دينار) لكل قسم كما هو موضح أدناه:

عدد العاملين	95	80	106
متوسط الأجر	110	107	99

يطلب حساب متوسط أجر العامل الشهري في هذا المصنع؟

4-8: جد الوسط الحسابي المرجح لمجموعات البيانات الآتية:

أ-

x_i	10	18	15	11	19	22	20
w_i	1	4	3	2	5	7	6

ب-

x_i	-1	0	2	-2	-3	1	-4
w_i	2	4	3	5	3	2	2

4-9: جد الوسط الحسابي المرجح لجدول التوزيع التكراري الآتي:

مراكز الفئات	التكرار	الأوزان
10	16	24
20	20	20
30	24	22
40	12	10
50	8	6
60	6	8

4-10: جد الوسط التوافقي والتربيعي لمجموعات البيانات الآتية:

أ-

10	14	12	18	16	20	25	19	15	17

ب-

-4	-3	-2	1	3	-2	5	-4	-3	2

ج-

0.2	0.5	0.3	0.1	0.4	0.8	0.6	0.7

4-11: جد الوسط (التوافقي، التربيعي والهندسي) لجدول التوزيع التكراري الآتي:

تسلسل الفئة	الفئات	التكرار
1	5-10	4
2	10-15	10
3	15-20	13
4	20-25	8
5	25-30	5

4-12: جد الوسط الهندسي لأسعار لتر البنزين في كل من مدن أقليم كوردستان:

أربيل	600	620	650	650	700	720
سليمانية	650	680	700	720	750	750
دهوك	580	600	590	650	680	700

4-13: لمجاميع البيانات الآتية جد المنوال:

A	4	6	8	6	7	5	5	7	6	6	7	8	6	9	6

B	1	2	3	4	5	6	7	8	9	10	11	12	13	14

C	2.5	3.5	3.25	3.3	2.52	3.4	3.29	3.6	2.5	3.4

4-14: فيما يلي تقديرات طلبة الإقتصاد في مادة الإحصاء مقابل عددهم:

ضعيف	مقبول	متوسط	جيد	جيد جداً	إمتياز
10	30	35	40	25	12

جد المنوال ؟

4-15: جد الوسيط لمجاميع البيانات الآتية:

21	5	17	18	19	20	8	7	4	22	12	10	3	9	7	2	A

ممتاز	ضعيف	متوسط	ممتاز	جيد	متوسط	جيد جدا	ضعيف	مقبول	B

4-16: جد الوسيط للبيانات المبوبة الآتية:

تسلسل الفئة	الفئات (كمية إنتاج الحنطة بالطن)	التكرار f_i (عدد قطع الأراضي)
1	26.70-17.83	1
2	35.56-26.70	14
3	44.42-35.56	20
4	53.28-44.42	8
5	62.15-53.28	7
6	71.02-62.15	6
7	79.88-71.02	4

الفصل الخامس

مقاييس التشتت

Measures of Variation

الفصل الخامس

مقاييس التشتت

Measures of Variation

1.5: مقدمة:

إن مقياس النزعة المركزية لايعطي وحده فكرة واضحة ومتكاملة عن مجموعة البيانات وخصوصاً فيما يتعلق الأمر بمقدار تجانس مفردات المجموعة مقارنةً بمجموعة أخرى من البيانات عن نفس الظاهرة، لذلك سيتم دراسة مقاييس التشتت والتي ستعطي فكرة واضحة عن صلاحية المتوسطات لتمثيل المفردات فإذا كانت نتيجة المقياس كبيرة كان التباعد بين قيم المجموعة واسعاً وعليه لايمكن الإعتماد على المتوسطات في تمثيل المفردات، أما إذا كانت النتيجة صغيرة فإن القيم متقاربة وتكون قريبة من المركز عندها يمكن الإعتماد على المتوسطات بالنسبة للمجموعات ذات المفردات القريبة من المتوسط.

2.5: المدى Range:

يعتبر المدى أبسط أنواع مقاييس التشتت المطلقة ويمكن حسابه في حالة البيانات المبوبة وفي حالة البيانات غير المبوبة وكما يلي:

أ- المدى للبيانات غير المبوبة:

يعرف المدى على أنه الفرق ما بين أكبر قيمة في مجموعة البيانات وأصغر قيمة فيها فإذا كانت $\left(x_1, x_2, \ldots, x_n\right)$ تمثل قياسات عينة من المفردات حجمها n وأن x_1 تمثل أكبر قيمة فيها وأن x_s تمثل أصغر قيمة فيها، عندئذٍ فإن المدى لهذه المجموعة هو:

$$\mathrm{R} = x_l - x_s \qquad \cdots \quad (5.1)$$

مثال (1.5)**: أفرض لديك البيانات التالية التي تمثل درجات عينة من الطلبة عددها 8 طلاب في مادة الإقتصاد، يطلب إيجاد مدى درجة الطالب في هذه العينة:**

x_i	65	66	90	56	75	43	78	25

الحل: أكبر درجة وأصغرها هي كما يأتي:

$$x_l = 90 \quad and \quad x_s = 25$$

لذلك فإن المدى هو: $R = x_l - x_s = 90 - 25 = 65$

ب- المدى للبيانات المبوبة:

يمكن إيجاد المدى للبيانات المبوبة من خلال حساب الفرق مابين مركز الفئة العليا والدنيا أو حساب الفرق مابين الحد الأعلى للفئة العليا والحد الأدنى للفئة الدنيا.

مثال (2.5)**: مصنع بيشار في أربيل لإنتاج البسكويت يتضمن 60 عاملاً سجلت أعدادهم مقابل إنتاجهم اليومي من خلال الجدول الآتي:**

فئات الإنتاج	30-24	37-31	44-38	51-45	58-52	65-59
عدد العمال	6	9	14	12	10	9

جد مدى إنتاجية العامل اليومية؟

الحل: يمكن حساب المدى من خلال حساب الفرق مابين مركز الفئة العليا والدنيا وكما يلي:

مركز الفئة الدنيا هو:

$$\frac{30 + 24}{2} = 27$$

مركز الفئة العليا هو:

$$\frac{65+59}{2}=62$$

إذن المدى هو: $R = 62 - 27 = 35$

أو باستخدام الفرق مابين الحد الأعلى للفئة العليا والحد الأدنى للفئة الدنيا. أي أن:

$$R = 65 - 24 = 41$$

ملاحظة: إن هذا المقياس سهل جداً ويستخدم في مجالات عديدة أهمها في تكوين لوحات السيطرة النوعية(Quality Control Charts) التي تستخدم في الرقابة على جودة الإنتاج وخصوصاً عندما يكون حجم العينة أقل من 10 مفردات، ولكن لايُعَوّل عليه كثيراً في قياس التشتت نظراً لكونه يعتمد على قيمتين فقط ويهمل بقية القيم الأخرى، لذلك فهو يتأثر كثيراً بالقيم المتطرفة ولكي يكون أكثر فائدة عادةً يهمل 5% من القيم المتطرفة الدنيا والعليا ومن ثم يتم حساب المدى الذي يعكس بصورة أفضل طبيعة توزيع البيانات في مجتمع الدراسة.

3.5: الإنحراف المتوسط Mean Deviation:

وهو عبارة عن متوسط إنحرافات قيم المجموعة عن وسطها الحسابي أي متوسط القيمة المطلقة لإنحرافات القيم عن وسطها الحسابي والذي يساوي صفر وللتخلص من ذلك تؤخذ القيمة المطلقة لإهمال الإشارة السالبة للإنحرافات، كما يمكن حسابه في حالة البيانات غير المبوبة وفي حالة البيانات المبوبة وكما يلي:

أ- الإنحراف المتوسط للبيانات غير المبوبة:

ويمكن حسابه من خلال الصيغة الآتية:

$$M.D=\frac{\sum_{i=1}^{n}\left|x_i-\bar{x}\right|}{n} \qquad \cdots \quad (5.2)$$

مثال (3.5): للبيانات التالية، جد الإنحراف المتوسط ؟

x_i	3	4	2	6	5

الحل:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{5} x_i}{5} = \frac{3+4+2+6+5}{5} = 4$$

فإن الإنحراف المتوسط هو كما يأتي:

$$M.D = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n} = \frac{|3-4| + |4-4| + |2-4| + |6-4| + |5-4|}{5} = \frac{6}{5} = 1.2$$

ب- الإنحراف المتوسط للبيانات المبوبة:

يتم حساب الإنحراف المتوسط للبيانات المبوبة من خلال إيجاد مراكز الفئات وضربها في التكرارات المناظرة لها لحساب الوسط الحسابي للبيانات المبوبة ومن ثم إيجاد إنحراف كل مركز فئة عن الوسط الحسابي مع إهمال الإشارة وضرب كل منها في التكرار المقابل له وجمعها وتقسيمها على مجموع التكرارات كما في الصيغة الآتية:

$$M.D = \frac{\sum_{i=1}^{m} f_i |x_i - \bar{x}|}{\sum_{i=1}^{m} f_i} \quad \cdots \quad (5.3)$$

وهنالك طرائق أخرى في حساب الإنحراف المتوسط من خلال إيجاد إنحرافات مراكز الفئات حول المنوال أو الوسيط بدلاً من الوسط الحسابي.

مثال (4.5)**: من خلال المثال** (2.5) **لجدول التوزيع التكراري التالي، جد الإنحراف المتوسط ؟**

فئات الإنتاج	30-24	37-31	44-38	51-45	58-52	65-59
عدد العمال	6	9	14	12	10	9

الحل: يمكن حساب الوسط الحسابي للبيانات المبوبة ومن ثم إيجاد الإنحراف المتوسط من خلال تكوين الجدول الآتي:

فئات الإنتاج	عدد العمال	x_i مراكز الفئات	$f_i x_i$	$\lvert x_i - \bar{x} \rvert$	$f_i \lvert x_i - \bar{x} \rvert$
30-24	6	27	162	18.4333	110.5998
37-31	9	34	306	11.4333	102.8997
44-38	14	41	574	4.4333	62.0662
51-45	12	48	576	2.5667	30.8004
58-52	10	55	550	9.5667	95.667
65-59	9	62	558	16.5667	149.1003
المجموع	60		2726		551.1334

$$\bar{x} = \frac{\sum_{i=1}^{m} f_i x_i}{\sum_{i=1}^{m} f_i} = \frac{2726}{60} = 45.4333$$

لذلك فإن الإنحراف المتوسط هو:

$$M.D = \frac{\sum_{i=1}^{m} f_i \lvert x_i - \bar{x} \rvert}{\sum_{i=1}^{m} f_i} = \frac{551.1334}{60} = 9.1856$$

ملاحظة: يعتبر هذا المقياس أدق من المدى لأنه يأخذ جميع القيم بنظر الإعتبار عند حسابه غير أنـه يتـأثر كثيراً بالقيم الشاذه وأن إهمال إشارة الإنحرافات تؤدي إلى محدودية الدقة.

4.5: الإنحراف الربيعي Quartile Deviation:

أو ما يسمى "نصف المدى الربيعي" والذي يعالج مشكلة التأثر بالقيم الـشاذه الـذي يعـاني منـه المدى ويمكن إيجاده في حالة البيانات المبوبة وفي حالة البيانات غير المبوبة وكما يلي:

أ- الإنحراف الربيعي للبيانات غير المبوبة:

يتم إيجاد الإنحراف الربيعي من خلال ترتيب البيانات تصاعدياً أو تنازلياً ومن ثم حساب الربيعين الأدنى والأعلى بمعرفة ترتيبهما وهما:

ترتيب الربيع الأدنى = $\frac{\text{عدد القيم} + 1}{4}$ ، ترتيب الربيع الأعلى = $\left(\frac{\text{عدد القيم} + 1}{4}\right) \times 3$

وإيجاد الإنحراف الربيعي من خلال الصيغة الآتية:

$$Q \cdot D = \frac{Q_u - Q_l}{2} \qquad \cdots \quad (5.4)$$

حيث أن Q_u تمثل الربيع الأعلى في حين يمثل Q_l الربيع الأدنى.

مثال (5.5)**: للبيانات التالية، جد الإنحراف الربيعي ؟**

x_i	4	3	2	6	5	2	3	5	1	0	4

الحل:

لإيجاد ترتيب الربيع الأدنى والأعلى نرتب البيانات تصاعدياً:

x_i	0	1	2	2	3	3	4	4	5	5	6

ترتيب الربيع الأدنى = $\frac{1 + 11}{4} = 3$

ترتيب الربيع الأعلى = $\left(\frac{1 + 11}{4}\right) \times 3 = 9$

أي أن: $Q_l = 2$ و $Q_u = 5$

يمكن حساب الإنحراف الربيعي كما يأتي:

$$Q \cdot D = \frac{Q_u - Q_l}{2} = \frac{5-2}{2} = 1.5$$

ب- الإنحراف الربيعي للبيانات المبوبة:

يتم إيجاد الإنحراف الربيعي للبيانات المبوبة من خلال تكوين التكرار المتجمع الصاعد (أو النازل) ومن ثم إيجاد الربيع الأدنى والأعلى وترتيبهم وكما يأتي:

ترتيب الربيع الأدنى = $\frac{\sum_{i=1}^{m} f_i}{4}$

ترتيب الربيع الأعلى = $3 \cdot \left(\frac{\sum_{i=1}^{m} f_i}{4} \right)$

قيمة الربيع الأدنى = الحد الأدنى للفئة التي يقع فيها الربيع الأدنى + $\frac{\text{ترتيب الربيع الأدنى - التكرار المتجمع الصاعد للفئة قبل الربيعية}}{\text{التكرار الأصلي لفئة الربيع الأدنى}}$ × طول فئة الربيع الأدنى

قيمة الربيع الأعلى = الحد الأدنى للفئة التي يقع فيها الربيع الأعلى + $\frac{\text{ترتيب الربيع الأعلى - التكرار المتجمع الصاعد للفئة قبل الربيعية}}{\text{التكرار الأصلي لفئة الربيع الأعلى}}$ × طول فئة الربيع الأعلى

على هذا الأساس يمكن أن نحصل على الإنحراف الربيعي من خلال إستخدام الصيغة (5.4)

مثال (6.5)**: لجدول التوزيع التكراري التالي، جد الإنحراف الربيعي؟**

فئات الدخل	عدد العوائل	الحدود العليا للفئات	التكرار المتجمع الصاعد
0-	18	200	18
200-	72	400	90
400-	154	600	244
600-	111	800	355
1000-800	45	1000	400

الحل: إيجاد الربيع الأدنى والأعلى وترتيبهم وكما يأتي:

ترتيب الربيع الأدنى = $\frac{\sum_{i=1}^{m} f_i}{4} = \frac{400}{4} = 100$

ترتيب الربيع الأعلى = $3 \cdot \left(\frac{\sum_{i=1}^{m} f_i}{4} \right) = 3 \cdot \left(\frac{400}{4} \right) = 300$

قيمة الربيع الأدنى = الحد الأدنى للفئة التي يقع فيها الربيع الأدنى + $\frac{\text{ترتيب الربيع الأدنى - التكرار المتجمع الصاعد للفئة قبل الربيعية}}{\text{التكرار الأصلي لفئة الربيع الأدنى}}$ × طول فئة الربيع الأدنى

قيمة الربيع الأدنى = $400 + \frac{100 - 90}{154} \times 200 = 412.987$ دينار

قيمة الربيع الأعلى = الحد الأدنى للفئة التي يقع فيها الربيع الأعلى + (ترتيب الربيع الأعلى - التكرار المتجمع الصاعد للفئة قبل الربيعية) / (التكرار الأصلي لفئة الربيع الأعلى) × طول فئة الربيع الأعلى

قيمة الربيع الأعلى = $600 + \frac{300 - 244}{111} \times 200 = 700.9009$ دينار

وعلى هذا الأساس يمكن الحصول على الإنحراف الربيعي وكما يأتي:

$$Q\cdot D = \frac{Q_u - Q_l}{2} = \frac{700.9009 - 412.987}{2} = 143.957$$

ملاحظة: يستخدم عادةً الإنحراف الربيعي عندما يكون هنالك قيم شاذه في مجموعة البيانات وهو أفضل من مقياس المدى ولكن يعاب عليه أنه مقياس تقريبي ولايأخذ جميع القيم بنظر الإعتبار عند حسابه.

5.5: الإنحراف المعياري Standard Deviation:

ويسمى في بعض الأحيان بالإنحراف القياسي ويعتبر هذا المقياس بحق أفضل مقاييس التشتت على الإطلاق لما يمتاز به من ميزات مثالية جعلته يقف في مقدمتها عند التطبيق، ويمكن إيجاده في حالة البيانات غير المبوبة وفي حالة البيانات المبوبة وكما يأتي:

أ- الإنحراف المعياري للبيانات غير المبوبة:

ويعرف على أنه الجذر التربيعي الموجب لمتوسط مجموع مربعات إنحرافات البيانات عن وسطها الحسابي. أي أن:

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \quad \cdots \quad (5.5)$$

وهنالك صيغ أخرى مشتقة من الصيغة(5.5) يمكن منها إيجاد الإنحراف المعياري وكما يأتي:

$$S = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}} \qquad \cdots \quad (5.6)$$

$$S = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1}} \qquad \cdots \quad (5.7)$$

مثال (7.5)**: البيانات التالية تمثل درجات قسم الإقتصاد في مادة الإحصاء وكما يأتي:**

x_i	50	70	54	45	60	65	85	38	90	83

والمطلوب حساب الإنحراف المعياري لدرجات الطلبة.

الحل: نجد أولاً الوسط الحسابي للبيانات أعلاه:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{50+70+\cdots+83}{10} = \frac{640}{10} = 64$$

ويمكن تسهيل الحسابات من خلال تكوين الجدول الآتي:

x_i^2	$(x_i - \bar{x})^2$	$x_i - \bar{x}$	البيانات x_i
2500	196	14-=64-50	50
4900	36	6=64-70	70
2916	100	10-=64-54	54

2025	361	19-=64-45	45
3600	16	4-=64-60	60
4225	1	1=64-65	65
7225	441	21=64-85	85
1444	676	26-=64-38	38
8100	676	26=64-90	90
6889	361	19=64-83	83
43824	2864	0	المجموع

من خلال الصيغة (5.5) يمكن الحصول على الإنحراف المعياري وكما يلي:

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{2864}{10-1}} = \sqrt{318.2222} = 17.8388$$

أو من خلال الصيغة (5.6) يمكن الحصول على الإنحراف المعياري وكما يلي:

$$S = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}} = \sqrt{\frac{(43824) - \frac{(640)^2}{10}}{10-1}}$$

$$= \sqrt{\frac{43824 - 40960}{9}} = \sqrt{\frac{2864}{9}} = \sqrt{318.2222} = 17.8388$$

أو من خلال الصيغة (5.7) يمكن الحصول على الإنحراف المعياري وكما يلي:

$$S = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2}{n-1}} = \sqrt{\frac{43824 - (10)(64)^2}{10-1}}$$

$$= \sqrt{\frac{43824 - (10)(4096)}{9}} = \sqrt{\frac{2864}{9}} = 17.8388$$

ب- الإنحراف المعياري للبيانات المبوبة:

يمكن حساب الإنحراف المعياري للبيانات المبوبة من خلال إيجاد مراكز الفئات وضربها في التكرار المقابل له ومنها يتم حساب الوسط الحسابي للبيانات المبوبة وإيجاد إنحرافات مراكز الفئات عن وسطها الحسابي وتطبيق الصيغة الآتية:

$$S = \sqrt{\frac{\sum_{i=1}^{m} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{m} f_i - 1}} \qquad \cdots \quad (5.8)$$

وهنالك عدة صيغ أخرى مشتقة من الصيغة أعلاه يمكن من خلالها حساب الإنحراف المعياري لامجال لذكرها في هذا الكتاب.

مثال (8.5)**: الجدول التالي يمثل فئات الإنتاج في معمل الأخوه شيشار وبيشار مقابل أعداد العاملين، والمطلوب إيجاد الإنحراف المعياري.**

الفئات (فئات الإنتاج)	التكرار (أعداد العاملين)
20-10	2
30-20	8
40-30	9
50-40	7
60-50	4

الحل: لحساب الإنحراف المعياري يتم تكوين الجدول الآتي:

الفئات	التكرار f_i	مراكز الفئات x_i	$f_i x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
20-10	2	15	30	21-	441	882
30-20	8	25	200	11-	121	968
40-30	9	35	315	1-	1	9
50-40	7	45	315	9	81	567
60-50	4	55	220	19	361	1444
المجموع	30		1080			3870

يتم أولاً حساب الوسط الحسابي وكما يأتي:

$$\bar{x} = \frac{\sum_{i=1}^{5} f_i x_i}{\sum_{i=1}^{5} f_i} = \frac{1080}{30} = 36$$

ومن خلال الصيغة (5.8) يتم حساب الإنحراف المعياري وكما يأتي:

$$S = \sqrt{\frac{\sum_{i=1}^{5} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{5} f_i - 1}} = \sqrt{\frac{3870}{30 - 1}} = \sqrt{133.4483} = 11.552$$

ملاحظة: يُعَدُ الإنحراف المعياري من أهم مقاييس التشتت على الإطلاق يقابل أهمية الوسط الحسابي لمقاييس النزعة المركزية وذلك كونه يأخذ جميع البيانات في نظر الإعتبار عند حسابه، ونظراً لدقته وسهولة حسابه جبرياً وخواصه الرياضية والإحصائية الممتازه التي تمكن استخدامه في التحليلات والقوانين التي تدخل درجة التشتت كعنصر فيها كما هو الحال في دراسة الإحتمالات والعينات

والإختبارات الإحصائية، ولكن من عيوبه تأثره بالقيم الشاذه ويصعب حسابه في البيانات الوصفية.

6.5: التباين The Variance:

التباين هو عبارة عن مربع الإنحراف المعياري لمجموعة من القيم وعليه فإن الرمز S^2 يـشير الى تباين قيم العينة، في حين أن σ^2 تشير إلى تباين المجتمع ككل ويمكن إيجاده في حالة البيانات غير المبوبة وفي حالة البيانات المبوبة وكما يأتي:

أ- التباين للبيانات غير المبوبة:

يمكن الحصول على التباين للبيانات غير المبوبة من خلال تربيع الإنحراف المعياري في الصيغة (5.5) أو (5.6) أو (5.7) أي من خلال حساب إنحرافات قيم مفردات العينة عن الوسط الحسابي مقسوماً على درجات الحرية (n-1) وبدون أخذ الجذر التربيعي، أي أن:

$$S^2 = \frac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}{n-1} \qquad \cdots \quad (5.9)$$

في حين يمكن حسابه للمجتمع من خلال الصيغة الآتية:

$$\sigma^2 = \frac{\sum_{i=1}^{N} \left(x_i - \bar{x}\right)^2}{N-1} \qquad \cdots \quad (5.10)$$

مثال (9.5): للبيانات الآتية، جد التباين ؟

x_i	5	7	4	5	6	5	6	2

الحل: نحسب أولاً الوسط الحسابي وكما يلي:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{8} x_i}{8} = \frac{5+7+\cdots+2}{8} = \frac{40}{8} = 5$$

ويمكن إيجاد تباين العينة كما يأتي:

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{8}(x_i - \bar{x})^2}{8-1} = \frac{(5-5)^2 + (7-5)^2 + \cdots + (2-5)^2}{7}$$

$$= \frac{0+4+\cdots+9}{7} = \frac{16}{7} = 2.2857$$

ب- التباين للبيانات المبوبة:

يمكن حساب تباين البيانات المبوبة من عينة من خلال تربيع الصيغة (5.8)، أي أن:

$$S^2 = \frac{\sum_{i=1}^{m} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{m} f_i - 1} \qquad \cdots \quad (5.11)$$

مثال (10.5)**: الجدول الآتي يمثل فئات الأجور مقابل عدد الموظفين في هيئة إحصاء أقليم كوردستان، والمطلوب تقدير تباين الأجور.**

الفئات (فئات الأجور)	التكرار (أعداد الموظفين)
300-200	20
400-300	30
500-400	60
600-500	44
700-600	32
800-700	14

الحل: لحساب التباين يتم تكوين الجدول الآتي:

الفئات	التكرار f_i	مراكز الفئات x_i	$f_i x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
300-200	20	250	5000	240-	57600	1152000
400-300	30	350	10500	140-	19600	588000
500-400	60	450	27000	40-	1600	96000
600-500	44	550	24200	60	3600	158400
700-600	32	650	20800	160	25600	819200
800-700	14	750	10500	260	67600	946400
المجموع	200		98000			3760000

يتم أولاً حساب الوسط الحسابي وكما يأتي:

$$\bar{x} = \frac{\sum_{i=1}^{6} f_i x_i}{\sum_{i=1}^{6} f_i} = \frac{98000}{200} = 490$$

ومن خلال الصيغة (5.11) يتم حساب التباين وكما يأتي:

$$S^2 = \frac{\sum_{i=1}^{6} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{6} f_i - 1} = \frac{3760000}{200 - 1} = 18894.4724$$

ملاحظة: إن ميزات وعيوب وخصائص التباين هي نفس ميزات وعيوب وخصائص الإنحراف المعياري، ويمكن تلخيص أهم هذه الخصائص بما يلي:

إن التباين دائماً أكبر أو يساوي الصفر.

1- إذا كانت $y_i = a x_i$ فإن $s_y^2 = a^2 \cdot s_x^2$

2- إذا كان $y_i = x_i \mp a$ فإن $s_y^2 = s_x^2$

3- إذا كان $y_i = a x_i \mp b$ فإن $s_y^2 = a^2 \cdot s_x^2$

7.5: معامل التشتت Coefficient of Dispersion:

أحياناً نحتاج إلى المقارنة بين تشتت مجموعتين أو أكثر من القيم المختلفة عن بعضها من حيث الوسط الحسابي، أو أن القيم للمجاميع المختلفة مقاسة بوحدات مختلفة فعندئذٍ لايمكن إستخدام مقاييس التشتت المطلق لوحده بل نحتاج إلى إستخدام مقاييس التشتت النسبية وهي خالية من وحدات القياس وهنالك عدة أنواع منها معامل التشتت المستند إلى المدى أو الأنحراف الربيعي أو الإنحراف المتوسط وأهمها الذي يستند إلى الإنحراف المعياري أو ما يسمى معامل الإختلاف (Coefficient of Variation) والذي يمكن إيجاده في حالة البيانات غير المبوبة وفي حالة البيانات المبوبة من خلال الصيغة الآتية:

$$C\cdot V = \frac{S}{\overline{x}} \cdot 100 \qquad \cdots \quad (5.12)$$

مثال (11.5)**: قارن بين التفاوت النسبي في الطول والوزن وبيّن أيهما أكثر تجانساً لمجموعة من الطلبة، إذا كان متوسط الطول 172 سم والإنحراف المعياري له 10.5 وكان متوسط الوزن 75 كغم والإنحراف المعياري له 15**

الحل: بما أن الظاهرتين مختلفتين من ناحية وحدة القياس لذلك يجب إستخدام مقياس تشتت نسبي مثل معامل الإختلاف للمقارنة بين مدى تفاوت التشتت بين الطول والوزن:

بالنسبة للطول: $$C\cdot V = \frac{S}{\overline{x}} \cdot 100 = \frac{10.5}{172} \cdot 100 = 6.1047$$

بالنسبة للوزن: $$C\cdot V = \frac{S}{\overline{x}} \cdot 100 = \frac{15}{75} \cdot 100 = 20$$

يتضح من ذلك أن الطلبة أكثر تجانساً في أطوالهم عما هو عليه في أوزانهم لأن معامل التشتت في الطول أقل من معامل التشتت في الوزن.

مثال (12.5)**: قارن بإستخدام معامل الإختلاف مدى التفاوت بين فئات الإنتاج في المثال** (8.5) **وفئات الأجور في المثال** (10.5).

الحل: من خلال المثال (8.5) حصلنا على ما يلي: الوسط الحسابي = 36 والإنحراف المعياري = 11.552 في حين بالنسبة للمثال (10.5) الوسط الحسابي = 490 والإنحراف المعياري = 18894.4724، لذلك يمكن إيجاد معامل الإختلاف كما يلي:

بالنسبة لفئات الإنتاج: $$C\cdot V = \frac{S}{\overline{x}} \cdot 100 = \frac{11.552}{36} \cdot 100 = 32.0889$$

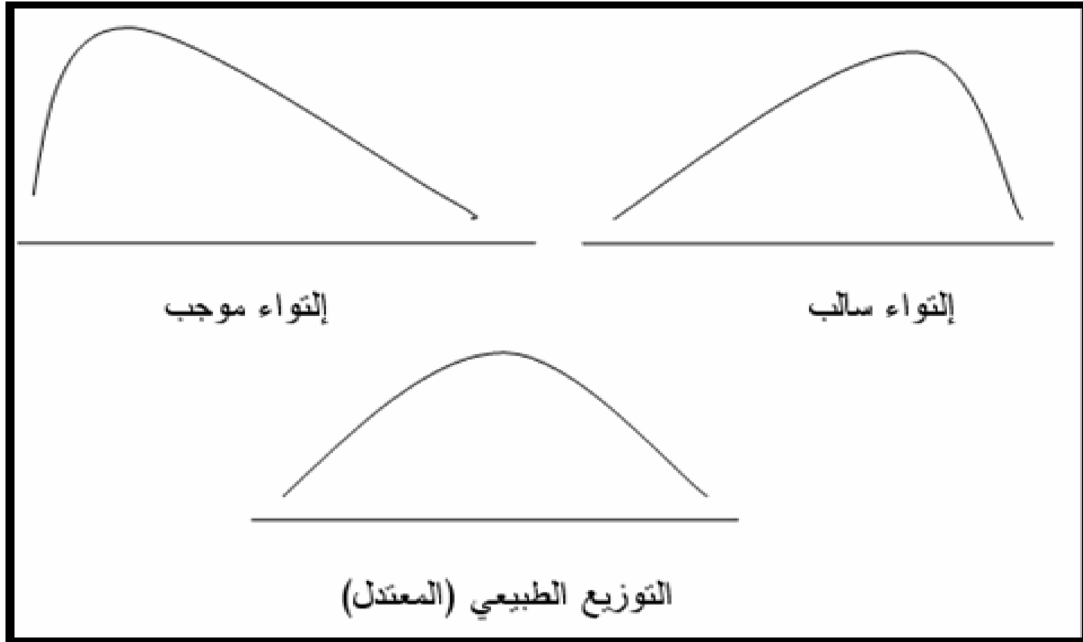
بالنسبة لفئات الأجور: $$C\cdot V = \frac{S}{\overline{x}} \cdot 100 = \frac{18894.4724}{490} \cdot 100 = 38.5601$$

يتضح من ذلك أن فئات الإنتاج أكثر تجانساً من فئات الأجور لأن معامل التشتت في فئات الإنتاج أقل من معامل التشتت في فئات الأجور.

8.5: الإلتواء والتفلطح Skew ness and Kurtosis:

المنحنيات التكرارية تكون متماثلة التوزيع أو غير متماثلة التوزيع، والمنحنيات غير المتماثلة تسمى (ملتوية) والإلتواء هو بعد المنحنى عن التماثل، وهذا يعني أن قيمته تعطي فكرة عن تمركز قيم البيانات فإذا كانت هذه البيانات تتمركز حول قيم البيانات الصغيرة أكثر من تمركزها حول قيم البيانات الكبيرة فإن توزيع هذه البيانات هو ملتوِ نحو اليمين ويسمى (موجب الإلتواء)، أما إذا كان العكس فإن إلتواء توزيع البيانات يكون سالباً أو ملتوياً نحو اليسار. أما إذا كان مقدار الإلتواء يساوي صفر فذلك يعني تماثل التوزيع.

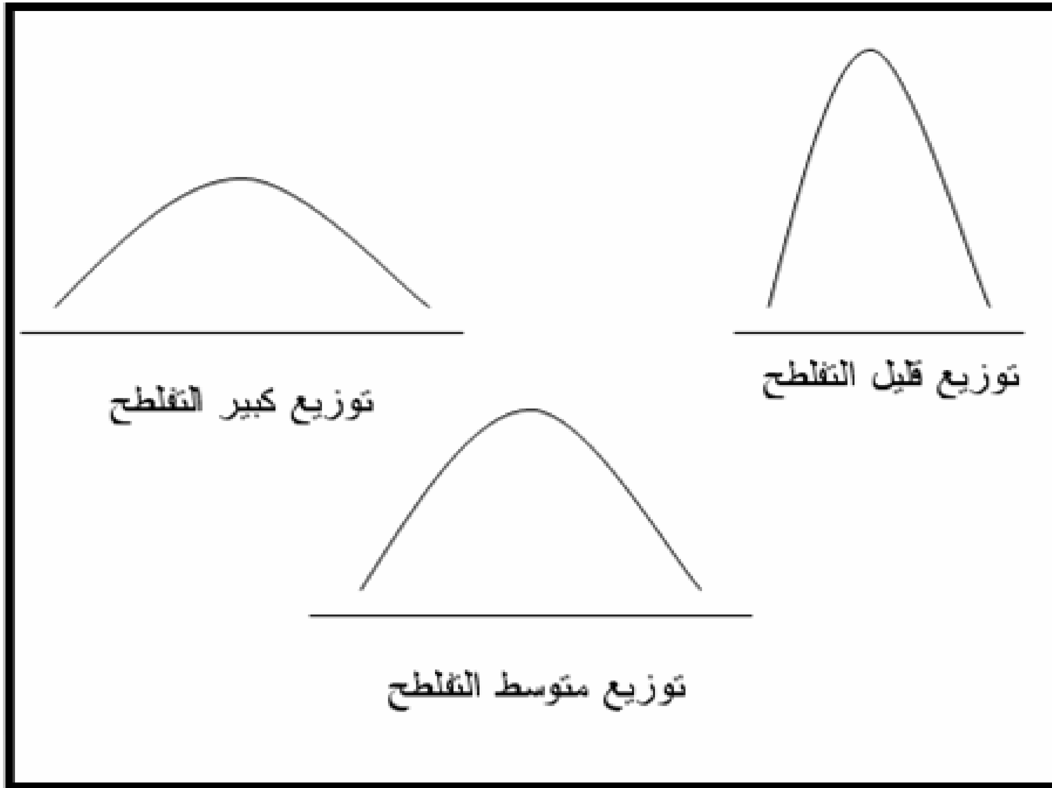
عندما يكون التوزيع ملتوياً نحو اليمين فإن القيم المتطرفة نحو اليمين تؤثر على الوسط الحسابي بسحبه نحو اليمين وبذلك يكون $\overline{x}$ أكبر من الوسيط أو المنوال. أما إذا كان التوزيع ملتوياً نحو اليسار فإن القيم المتطرفة الصغيرة تسحبه إلى اليسار ولذلك يكون الوسط الحسابي أصغر من الوسيط أو المنوال، أما إذا كان التوزيع معتدلاً فإن الوسط الحسابي مساوياً للوسيط أو المنوال. والأشكال التالية توضح ذلك:



الشكل (5.1)

يبين التوزيع المتماثل والملتوي السالب والموجب

التفلطح أو التفرطح يمثل تكرارات القيم على طرفي هذا التوزيع، وهو يمثل أيضاً درجة علو قمة التوزيع بالنسبة للتوزيع الطبيعي، فإذا كانت قيمة التفلطح كبيرة كانت للتوزيع قمة منخفضة ، ويسمى التوزيع كبير التفلطح، أما إذا كانت قيمة التفلطح صغيرة فإن للتوزيع قمة عالية ويسمى التوزيع مدبباً أو قليل التفلطح وإذا كانت قيمة التفلطح متوسطة يسمى التوزيع متوسط التفلطح. والأشكال التالية توضح ذلك:



الشكل (5.2)

يبين التوزيع متوسط وقليل وكبير التفلطح

وهنالك طرائق عديدة لحساب الإلتواء والتفلطح لامجال لذكرها في هذا الكتاب.

تمارين الفصل الخامس

1-5: ما المقصود بمقاييس التشتت ؟

2-5: جد المدى والإنحراف المتوسط للإنفاق الشهري لعينة مأخوذه من أربيل وكما يلي:

520	740	950	650	500	200	900	1000	800	700	400

3-5: إذا كان الإستهلاك العائلي الشهري للأرز في مدينة أربيل والسليمانية كما يلي:

اربيل	7.8	7.50	6.5	5.9	2.70	9.06	10.3	13.9	12.4	15	12
السليمانية	3.4	4.9	13	12	5.8	7.9	4.8	6.8	5.9	10	9.8

والمطلوب معرفة أي من المدينتين أكثر تجانساً بالإعتماد على المدى والإنحراف المتوسط.

4-5: البيانات التالية تمثل أوزان عينة من الطلبة عددها 15 طلاب:

x_i	45	57	46	53	48	60	59
67	62	54	49	44	68	55	46

يطلب حساب الإنحراف الربيعي والمعياري والتباين ؟

5-5: جد الإنحراف الربيعي والمعياري والتباين لجدول التوزيع التكراري الآتي:

الفئات	2-0	4-2	6-4	8-6	10-8
التكرار	4	6	8	6	4

6-5: قارن بين الجدوليين التكراريين الآتيين باستخدام معامل الإختلاف

الفئات	6-2	10-6	14-10	18-14	22-18
التكرار	5	7	8	4	3
الفئات	3-0	6-3	9-6	12-9	15-12
التكرار	14	16	18	16	14



الفصل السادس

تحليل الإرتباط

Correlation Analysis

الفصل السادس

تحليل الإرتباط

Correlation Analysis

1.6: مقدمة:

إن ما يطلق عليه بالإرتباط أو التلازم (Correlation) عبارة عن أداة تحليل إحصائي لقياس مدى الترابط وإتجاهاته بين متغيرين أو أكثر، وقياس الترابط هو لمعرفة طبيعة التغير الحاصل بين المتغيرين أو أكثر بحكم صفات معينة تجمع بينهما، فالمطلوب معرفته في هذه الحالة إذن هو تقدير ما إذا كانت صفات أحد المتغيرات متصلة ومرتبطة بصفات متغير آخر أو متغيرات أخرى والأمثلة على ذلك كثيرة، فمثلاً دراسة العلاقة بين الدخل الشهري والإنفاق الشهري للإسرة أي أنه يحدث تذبذب في الإنفاق الشهري مرتبطاً بتذبذب الدخل الشهري للأسرة، أي أن الملاحظ هو أن الإنفاق يزداد عادة عند زيادة الدخل الشهري للأسرة، كذلك يمكن ملاحظة العلاقة بين أسعار السلع والكمية المعروضة منها في السوق. ويجب أن يكون هنالك أولاً علاقة منطقية وعلمية بين متغيرين ثم يأتي عملية حساب الإرتباط بينهما فمثلاً ليس هنالك علاقة منطقية بين سرعة المياه في الوديان وأسعار الأراضي في المدينة، والتي لايمكن إيجاد إرتباط لعدم وجود علاقة منطقية (علمية) بينهما. وهنالك أنواع عديدة من معامل الإرتباط سنتناول في هذا الفصل أهمها.

2.6: معامل الإرتباط الخطي البسيط Simple Linear Correlation Coefficient:

أو مايسمى بمعامل إرتباط (Pearson) ويستخدم لدراسة العلاقة بين متغيرين لهما خصائص قابلة للقياس مثل الوزن، الطول، العمر، الدينار،...الخ ويرمز له للعينة r_{xy} في حين للمجتمع ρ_{xy}، وتوجد عدة صيغ رياضية لحساب معامل الإرتباط للبيانات المرتبة على شكل أزواج بالشكل الآتي: $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ حيث

يتم حساب معامل الإرتباط الخطي البسيط بإستخدام الصيغة الآتية:

$$r_{xy} = \frac{\text{cov}(x, y)}{S_x . S_y} = \frac{S_{xy}}{S_x . S_y} \quad \cdots \quad (6.1)$$

حيـث أن $\text{cov}(x, y) = S_{xy}$ يـسمى التبـاين المـشترك Covariance بـين قـيم x و y ويـتم حسابه وفق الصيغة الآتية:

$$\text{cov}(x, y) = S_{xy} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{n - 1} \quad \cdots \quad (6.2)$$

وعليه فإن:

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) \Big/ n - 1}{\sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n - 1}} . \sqrt{\frac{\sum_{i=1}^{n} (y_i - \overline{y})^2}{n - 1}}}$$

$$\therefore r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2 . \sum_{i=1}^{n} (y_i - \overline{y})^2}} \quad \cdots \quad (6.3)$$

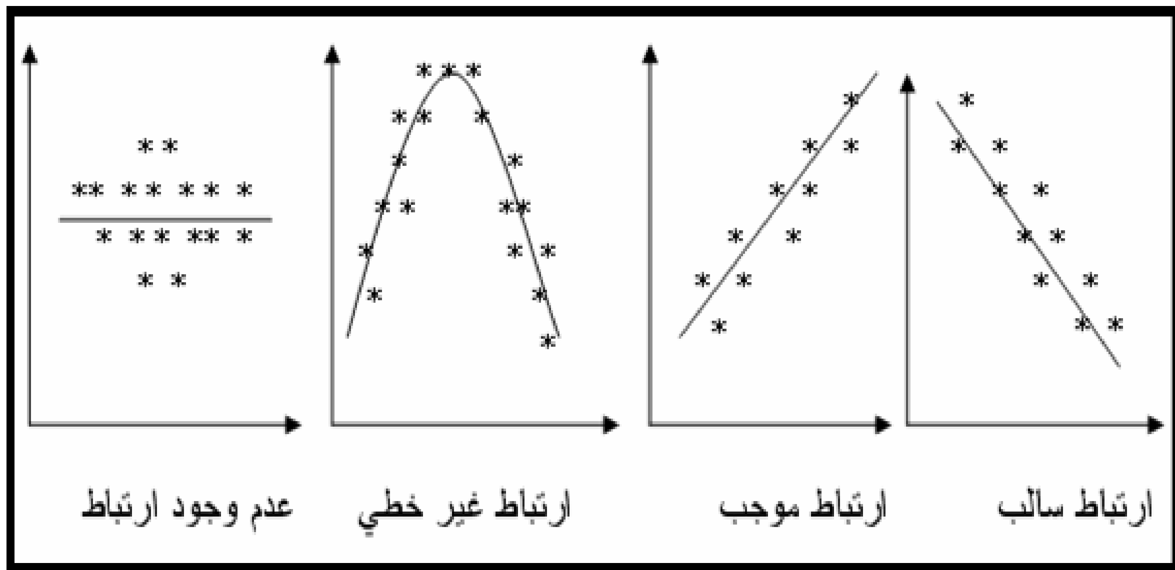
نلاحظ من خلال الـصيغة أعـلاه أن المقـام كميـة موجبـة دائمـا، وهـذا يعنـي أن إشـارة معامـل الإرتباط تتحدد من خلال إشارة البسط فإذا كان البسط (التباين المشترك) موجباً فذلك يعنـي أن الإرتبـاط موجب والعكس صحيح أيضاً، وتكون قيمته كما يأتي:

$$-1 \le \rho_{xy} \text{ أو } r_{xy} \le 1$$

كذلك لدينا:



الشكل الإنتشاري لأزواج القيم يوضح هل هنالك علاقة خطية (سالبة أو موجبـــة) أو غيـــر خطية أو عدم وجود علاقة بين المتغيرين والشكل الآتي يوضح ذلك :



الشكل (6.1)

يبين أشكال إنتشارية لبعض حالات الإرتباط وعدمه

مثال (1.6)**: أراد باحث إقتصادي دراسة العلاقة بين الدخل الشهري لـلأسرة والإنفـاق الـشهري لـلإسرة في مدينة السليمانية من خلال عينة عشوائية وكانت كما يلي:**

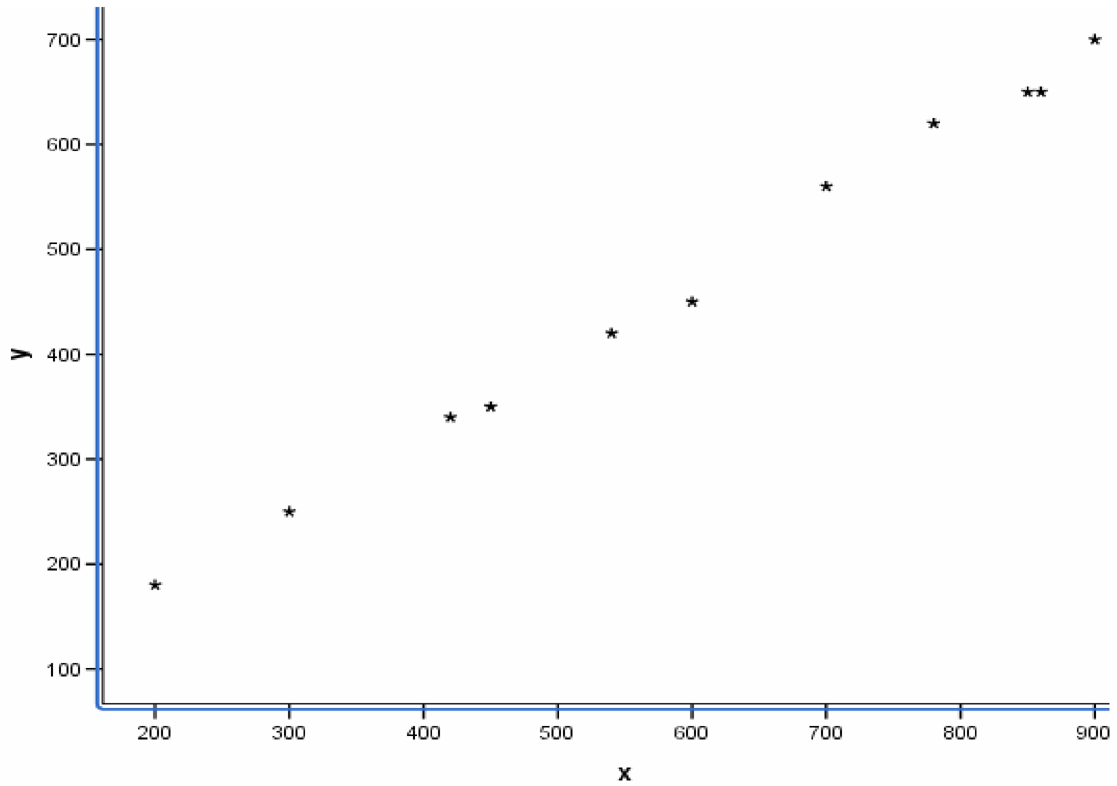
الدخل الشهري للأسرة x (آلاف الدنانير)	الإنفاق الشهري للأسرة y (آلاف الدنانير)
420	340
600	450
200	180
700	560
900	700
850	650
300	250
540	420
780	620
860	650
450	350

والمطلوب مايلي:

1- **أرسم شكل الإنتشار للبيانات أعلاه.**

2- **ما هو مقدار الإرتباط بين هاتين الظاهرتين.**

الحل: 1- يمكن رسم الإنتشار من خلال أزواج القيم x و y **وكما يأتي:**



الشكل (6.2)

يبين الشكل الإنتشاري لأزواج القيم x و y

من خلال الشكل الإنتشاري نلاحظ هنالك علاقة طردية خطية بين الدخل والإنفاق الشهري للأسرة في مدينة السليمانية.

2- لإيجاد معامل الإرتباط الخطي يتم حساب الأوساط الحسابية وتكوين الجدول التالي:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{11} x_i}{11} = \frac{420+600+\cdots+450}{11} = \frac{6600}{11} = 600$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{11} y_i}{11} = \frac{340+450+\cdots+350}{11} = \frac{5170}{11} = 470$$

$(y_i-\bar{y})^2$	$(x_i-\bar{x})^2$	$(x_i-\bar{x})(y_i-\bar{y})$	$(y_i-\bar{y})$	$(x_i-\bar{x})$	y	x	التسلسل
16900	32400	23400	130-	180-	340	420	1
400	0	0	20-	0	450	600	2
84100	160000	116000	290-	400-	180	200	3
8100	10000	9000	90	100	560	700	4
52900	90000	69000	230	300	700	900	5
32400	62500	45000	180	250	650	850	6
48400	90000	66000	220-	300-	250	300	7
2500	3600	3000	50-	60-	420	540	8
22500	32400	27000	150	180	620	780	9
32400	67600	46800	180	260	650	860	10
14400	22500	18000	120-	150-	350	450	11
315000	571000	423200	0	0	5170	6600	المجموع

ومن خلال الصيغة (6.3) يمكن الحصول على الإرتباط الخطي البسيط وكما يأتي:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 . \sum_{i=1}^{n}(y_i-\bar{y})^2}} = \frac{423200}{\sqrt{(571000)(315000)}}$$

$$= \frac{423200}{424104.9398} = 0.9978$$

وهذا يعني أن هنالك إرتباط طردي (موجب) قوي يساوي99.78% بين الـدخل والإنفـاق الـشهري للعائلة في مدينة السليمانية.

ويمكن إيجاد الإرتباط الخطي البسيط بطريقة أخرى وذلك من خلال حـساب الإنحـراف المعيـاري لكل من الدخل والإنفاق ومن ثم حساب التباين المشترك وكما يلي:

$$S_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{571000}{11-1}} = \sqrt{57100} = 238.9561$$

$$S_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{315000}{11-1}} = \sqrt{31500} = 177.4824$$

$$\operatorname{cov}(x,y) = S_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{423200}{11\text{-}1} = 42320$$

ومن خلال الصيغة (6.1) يمكن الحصول على معامل الإرتباط الخطي البسيط وكما يأتي:

$$r_{xy} = \frac{\operatorname{cov}(x,y)}{S_x . S_y} = \frac{S_{xy}}{S_x . S_y}$$

$$= \frac{42320}{(238.9561)(177.4824)} = 0.9978$$

مثال :(2.6)البيانات التالية تمثل الكمية المطلوبة من سلعة معينة وسعر الوحدة الواحدة من هذه السلعة:

الكمية المطلوبة y	9	5	7	8	9	10
السعر x	5	7	6	5	4	3

يطلب حساب معامل الإرتباط الخطي البسيط بين الكمية المطلوبة والسعر.

الحل:

$$\bar{x}=\frac{\sum_{i=1}^{n}x_i}{n}=5 \quad , \quad \bar{y}=8$$

لحساب معامل الإرتباط الخطي البسيط يتم تكوين الجدول الآتي:

$(y_i-\bar{y})^2$	$(x_i-\bar{x})^2$	$(x_i-\bar{x})(y_i-\bar{y})$	$y_i-\bar{y}$	$x_i-\bar{x}$	y_i	x_i
1	0	0	1	0	9	5
9	4	-6	-3	2	5	7
1	1	-1	-1	1	7	6
0	0	0	0	0	8	5
1	1	-1	1	-1	9	4
4	4	-4	2	-2	10	3
16	10	-12	0	0	المجموع	

وعليه فإن:

$$r_{xy}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2.\sum_{i=1}^{n}(y_i-\bar{y})^2}}=\frac{-12}{\sqrt{10}.\sqrt{16}}$$

$$=\frac{-12}{(3.1623)(4)}=\frac{-12}{26.533}=-0.9487$$

وهذا يعني أن درجة الإرتباط الخطي البسيط مابين الكمية المطلوبة من هذه السلعة وسعر الوحدة منها هو ‎-94.87%‎ وأنه إرتباط عكسي (سالب) (دلالة على أنه كلما ارتفع السعر إنخفضت بالمقابل الكمية المطلوبة منها) على ضوء هذه البيانات.

3.6: معامـل إرتبـاط الرتـب لـسبيرمان Spearman's rank Correlation Coefficient:

أحياناً تكون البيانات من النوع الوصفي أو كبيرة جداً والمطلوب قياس قوة الإرتباط الخطي ففي هذه الحالة يمكن استخدام الرتب وذلك من خلال تحديد رتب الصفات (أو الأرقام الكبيرة) المراد دراسـتها مثل تقديرات الطالب في الإمتحانات النهائية، المستوى المعيشي، التحـصيل العلمـي، كفـاءة العامـل،...الـخ.

ويمكن استخدام هذا الإرتباط عندما يكون أحد أو كلا الظاهرتين عبارة عن بيانات كمية.

إرتباط الرتب لسبيرمان يمكن استخدامه من خلال ترتيب الصفات (الأرقام الكبـيرة) تـصاعدياً أو تنازلياً وإعطاء كل صفة من هذه الصفات قيمة عددية متسلسلة من 1 إلى n وبذلك تحولـت الـصفات إلى قيم عددية، ففي حالة عدم وجود تكرارات في تلك الصفات نستخدم الصيغة الآتية:

$$r_{xy} = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{n(n^2-1)} \qquad \cdots \quad (6.4)$$

حيـث تمثـل d_i الفـرق بـين رتـب المتغـيرين، أي أن: $d_i = rx_i - ry_i$ ويجـب أن يـساوي مجموعه الصفر أي أن: $\sum_{i=1}^{n} d_i = 0$

مثال (3.6): البيانات التالية تمثل تقـديرات المـستوى المعيـشي مقابـل التحـصيل العلمـي لعـدة أشـخاص، والمطلوب حساب الإرتباط الخطي البسيط بينهما.

المستوى المعيشي x	جيد	ممتاز	سيىء	جيد جداً	متوسط	مقبول
التحصيل العلمي y	إعدادية	ماجستير	يقرأ ويكتب	بكالوريوس	إبتدائية	أمي

الحل: يتم أولاً ترتيب صفات كل متغير تصاعدياً وإعطاء قيمة عددية لكل منها وكما يلي:

ترتيب المستوى المعيشي x تصاعدياً	سيىء	مقبول	متوسط	جيد	جيد جداً	ممتاز
الرتبة rx_i	1	2	3	4	5	6

ترتيب التحصيل العلمي y تصاعدياً	أمي	يقرأ ويكتب	إبتدائية	إعدادية	بكالوريوس	ماجستير
الرتبة ry_i	1	2	3	4	5	6

وعلى هذا الأساس يمكن تكوين الجدول الآتي:

المستوى المعيشي x	التحصيل العلمي y	الرتبة rx_i	الرتبة ry_i	d_i	d_i^2
جيد	إعدادية	4	4	0	0
ممتاز	ماجستير	6	6	0	0
سيىء	يقرأ ويكتب	1	2	-1	1
جيد جداً	بكالوريوس	5	5	0	0
متوسط	إبتدائية	3	3	0	0
مقبول	أمي	2	1	1	1
المجموع				0	2

ومن خلال الصيغة (6.4) يمكن الحصول على الإرتباط الخطي البسيط باستخدام طريقة سبيرمان وكما يلي:

$$r_{xy} = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 1 - \frac{6 \cdot (2)}{6 \cdot (6^2 - 1)}$$

$$= 1 - \frac{12}{6 \cdot 35} = 1 - \frac{12}{210} = 1 - 0.0571 = 0.9429$$

وهـذا يعنـي أن هنالـك علاقـة طرديـة (موجبـة) وقويـة تـساوي 94.29% بـين المـستوى المعيـشي والتحصيل الدراسي.

أما في حالة وجود تكرار في صفات العينة لأحد المتغيرين أو كلاهما فذلك يـستوجب إجـراء بعـض التعديلات على طريقة سبيرمان وكما يأتي:

يتم أولاً ترتيب الصفات تنازلياً أو تصاعدياً ومن ثم يخـصص لكـل صـفة قيمـة عدديـة حتـى وأن تكررت في العينة ومن ثم يتم حساب معدل القيم المخصصة وإعادة تخصيصه لتلك الصفات المكررة، ومن ثم يتم إجراء تعديل من خلال إضـافة الكميـة k إلى $\sum_{i=1}^{n} d_i^2$ في الـصيغة الـسابقة للحـصول عـلى الإرتبـاط الخطي البسيط، أي أن:

$$r_{xy} = 1 - \frac{6 \cdot \left(\sum_{i=1}^{n} d_i^2 + k \right)}{n(n^2-1)} \qquad \cdots \quad (6.5)$$

حيث تمثل k مجموع القيم المكررة التي نحصل عليها من خلال الصيغة الآتية:

$$k = \sum_{i=1}^{u} \frac{m_i (m_i^2 - 1)}{12} \qquad \cdots \quad (6.6)$$

u تمثل عدد حالات وجود صفات مكررة في المتغيرين، وتمثل m_i عدد مرات تكرار الصفة i.

مثال (4.6): البيانات التالية تمثل تقديرات مجموعـة مـن طلبـة قـسم الإدارة في مـادتي الإدارة والإحـصاء، والمطلوب حساب معامل إرتباط سبيرمان.

الإدارة x	ممتاز	جيد	ممتاز	ضعيف	جيد جداً	متوسط	مقبول
الإحصاء y	جيد	مقبول	جيد جداً	مقبول	جيد	جيد	متوسط

الحل: يتم أولاً ترتيب صفات كل متغير تصاعدياً وإعطاء قيمة عددية لكل منها وكما يلي:

ترتيب تقديرات مادة الإدارة	ضعيف	مقبول	متوسط	جيد	جيد جداً	ممتاز	ممتاز
الرتبة rx_i	1	2	3	4	5	6	7
المعدل						6.5	

ترتيب تقديرات مادة الإحصاء	مقبول	مقبول	متوسط	جيد	جيد	جيد	جيد جداً
الرتبة ry_i	1	2	3	4	5	6	7
المعدل	1.5			5			

من خلال البيانات أعلاه نلاحظ وجـود ثـلاث حـالات تكـرار u = 3 وأن الحالـة الأولى المكـررة هـي ممتاز وتكررت مرتان $m_1 = 2$ وأن الحالة الثانية المكررة هي مقبول وقد تكررت مرتـان $m_2 = 2$ وأن الحالة الثالثة المكررة هي جيد وقد تكررت ثلاث مرات $m_3 = 3$ لذلك فإن قيمـة k يمكـن حـسابها كما يأتي:

$$k = \sum_{i=1}^{3} \frac{m_i\left(m_i^2 - 1\right)}{12} = \frac{m_1\left(m_1^2 - 1\right)}{12} + \frac{m_2\left(m_2^2 - 1\right)}{12} + \frac{m_3\left(m_3^2 - 1\right)}{12}$$

و

$$= \frac{2\left(2^2 - 1\right)}{12} + \frac{2\left(2^2 - 1\right)}{12} + \frac{3\left(3^2 - 1\right)}{12} = 0.5 + 0.5 + 2 = 3$$

يمكن تكوين الجدول الآتي:

الإدارة x	الإحصاء y	الرتبة rx_i	الرتبة ry_i	d_i	d_i^2
ممتاز	جيد	6.5	5	1.5	2.25
جيد	مقبول	4	1.5	2.5	6.25
ممتاز	جيد جداً	6.5	7	-0.5	0.25
ضعيف	مقبول	1	1.5	-0.5	0.25
جيد جداً	جيد	5	5	0	0
متوسط	جيد	3	5	-2	4
مقبول	متوسط	2	3	-1	1
المجموع				0	14

وباستخدام الصيغة (6.5) نحصل على ما ياتي:

$$r_{xy} = 1 - \frac{6 \cdot \left(\sum_{i=1}^{n} d_i^2 + k \right)}{n(n^2-1)} = 1 - \frac{6.(14+3)}{7(7^2-1)} = 1 - \frac{102}{336} = 1 - 0.3036 = 0.6964$$

وهذا يعني هنالك علاقة طردية (موجبة) وقوية بين تقديرات مادتي الإدارة والإحصاء.

4.6: معامل الإرتباط الجزئي Partial Correlation Coefficient:

هو دراسة العلاقة بين متغيرين (x_1 و x_2 مثلاً) بإستبعاد أثر x_3 على كلا المتغيرين وأننا نرغب في قياس درجة الإرتباط بين x_1 و x_2 بإستبعاد أثر المتغير الثالث x_3 على كلا المتغيرين ويرمز له عادة $r_{12.3}$. ويمكن الحصول عليه من خلال الصيغة الآتية:

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}} \qquad \cdots \quad (6.7)$$

حيث تمثل r_{12} ، r_{13} و r_{23} الإرتباط الخطي البسيط بين متغيرين التي يمكن أن نحصل عليها من خلال الصيغة (6.3)، وبنفس الطريقة يمكن الحصول على $r_{13.2}$ أي الإرتباط الجزئي بين x_1 و x_3 بإستبعاد أثر المتغير الثاني x_2 وهكذا بالنسبة إلى $r_{23.1}$.

أما إذا كانت الدراسة تحتوي على دراسة العلاقة بين متغيرين x_1 و x_2 بإستبعاد أثـر المتغـيرين x_3 و x_4 فيرمز له $r_{12.34}$ ويمكن الحصول عليه من خلال الصيغة الآتية:

$$r_{12.34} = \frac{r_{12.3} - r_{14.3}r_{24.3}}{\sqrt{(1 - r_{14.3}^2)(1 - r_{24.3}^2)}} \qquad \cdots \quad (6.8)$$

حيث تمثل $r_{12.3}$ ، $r_{14.3}$ $r_{24.3}$ الإرتباط الجزئي الذي يمكن أن نحصل عليـه مـن خلال الصيغة (6.7) وهكذا... الى إستبعاد k من المتغيرات أي $r_{12.34...k}$.

مثال (5.6): أفرض أن x_1 يمثل الإنفاق الشهري للأسرة و x_2 يمثل دخلها الشهري و x_3 يمثل عدد أفراد الأسرة، وعلى أساس عينة عشوائية من الأسر قوامها (8) أسر تم الحصول على البيانات الآتية:

الإنفاق الشهري x_1	الدخل الشهري x_2	عدد أفراد الأسرة x_3
300	400	3
700	900	5
250	500	2
850	950	6
650	800	5
540	600	4
200	300	2
450	500	3

المطلوب ما يلي:

1- **إيجاد الإرتباط الجزئي بين** x_1 **و** x_2 **بإستبعاد أثر المتغير** x_3

2- **إيجاد الإرتباط الجزئي بين** x_2 **و** x_3 **بإستبعاد أثر المتغير** x_1

3- **إيجاد الإرتباط الجزئي بين** x_1 **و** x_3 **بإستبعاد أثر المتغير** x_2

الحل: لإيجاد الإرتباط الجزئي بين متغيرين بإستبعاد أثر المتغير الثالث يجب أولاً حساب الإرتباط الخطي البسيط بين كل متغيرين أي r_{12} ، r_{13} و r_{23} وكما يلي:

$$r_{12} = \frac{s_{x_1 x_2}}{s_{x_1} s_{x_2}} = \frac{53360.7622}{(233.3452)(238.9523)} = 0.957$$

$$r_{13} = \frac{s_{x_1 x_3}}{s_{x_1} s_{x_3}} = \frac{340.6205}{(233.3452)(1.488)} = 0.981$$

$$r_{23} = \frac{s_{x_2 x_3}}{s_{x_2} s_{x_3}} = \frac{333.8718}{(238.9523)(1.488)} = 0.939$$

1- لإيجاد الإرتباط الجزئي بين x_1 و x_2 بإستبعاد أثر المتغير x_3 لدينا مايلي:

$$r_{12.3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}} = \frac{0.957 - (0.981)\cdot(0.939)}{\sqrt{(1 - 0.981^2)(1 - 0.939^2)}}$$

$$= \frac{0.0358}{\sqrt{(0.0376)(0.1183)}} = \frac{0.0358}{0.0667} = 0.5368$$

وهذا يعني أن هنالك إرتباط طردي قوي يساوي 53.68% بين الإنفاق الشهري والدخل الشهري للأسرة بإستبعاد أثر متغير عدد أفراد الأسرة.

2- لإيجاد الإرتباط الجزئي بين x_2 و x_3 بإستبعاد أثر المتغير x_1 لدينا ما يلي:

$$r_{23.1} = \frac{r_{23} - r_{12} r_{13}}{\sqrt{(1 - r_{12}^2)(1 - r_{13}^2)}} = \frac{0.939 - (0.957)\cdot(0.981)}{\sqrt{(1 - 0.957^2)(1 - 0.981^2)}}$$

$$\therefore \quad r_{23.1} = \frac{0.000183}{\sqrt{(0.0842)(0.0376)}} = \frac{0.000183}{0.0563} = 0.0033$$

وهذا يعني أن هنالك إرتباط طردي ضعيف جداً يساوي 0.33% بين الدخل الشهري للأسرة وعدد أفراد الأسرة بإستبعاد أثر متغير الإنفاق الشهري للأسرة.

3- لإيجاد الإرتباط الجزئي بين x_1 و x_3 بإستبعاد أثر المتغير x_2 لدينا ما يلي:

$$r_{13.2} = \frac{r_{13} - r_{12} r_{23}}{\sqrt{(1 - r_{12}^2)(1 - r_{23}^2)}} = \frac{0.981 - (0.957)\cdot(0.939)}{\sqrt{(1 - 0.957^2)(1 - 0.939^2)}}$$

$$\therefore \quad r_{13.2} = \frac{0.0824}{\sqrt{(0.0842)(0.1183)}} = \frac{0.0824}{0.0998} = 0.8257$$

وهذا يعني أن هنالك إرتباط طردي قوي يساوي 82.57% بين الإنفاق الـشهري وعـدد أفـراد الأسرة بإستبعاد أثر متغير الدخل الشري للأسرة.

مثال (6.6): جد الإرتباط الجزئي بـين x_1 و x_2 بإسـتبعاد أثـر المتغـيرين x_3 و x_4 إذا تـوفرت لـديك المعلومات الآتية:

$r_{12.3} = 0.903$	$r_{14.3} = 0.302$	$r_{24.3} = 0.216$

الحل:

$$r_{12.34} = \frac{r_{12.3} - r_{14.3}r_{24.3}}{\sqrt{(1 - r_{14.3}^2)(1 - r_{24.3}^2)}}$$

$$= \frac{0.903 - (0.302)(0.216)}{\sqrt{(1 - 0.302^2)(1 - 0.216^2)}}$$

$$= \frac{0.8378}{\sqrt{(0.9088)(0.9533)}} = \frac{0.8378}{0.9308} = 0.90$$

وهذا يعني أن هنالك إرتباط طردي قوي يساوي 90% بين x_1 و x_2 بإستبعاد أثر المتغـيرين x_3 و x_4.

5.6: معامل الإرتباط المتعدد :Multiple Correlation Coefficient في كثـير مـن الأحيـان نلاحـظ أن التغير الذي يحدث في ظاهرة ما ربما يكون بسبب تغير مجموعة من الظواهر الأخرى مجتمعـةً ولـيس بسبب ظاهرة واحدة فقط مثل ما كان عليه الحال عند دراسـة معامـل الإرتبـاط الخطـي البـسيط. وعـلى هذا الأساس فإن التغير في الإنفاق قد يكـون سـببه تغـير الـدخل فـضلاً عـن تغـير عـدد أفـراد الأسرة، وأن التغير في إنتاجية الدونم الواحد من الحنطة أو الشعير قد يتأثر بتغير نوع البذور المستخدمة وتغير كميات الأمطار الموسمية وتغير نسبة الملوحـة في التربـة،...الـخ. لـذلك يمكـن تعريـف الإرتبـاط المتعـدد عـلى أنـه عبــارة عــن قيــاس قــوة العلاقــة بــين متغــير مــن جهــة مــع عــدة متغــيرات (أكــثر مــن متغــير

واحد) من جهة أخرى ويرمز لها مثلاً $r_{1.23}$ التي تعني الإرتباط المتعدد لعينة بين المتغير x_1 مع كلا المتغيرين الثاني والثالث مجتمعان أو $r_{1.234}$ التي تعني الإرتباط المتعدد لعينة بين المتغير x_1 مع المتغيرات الثلاث الأخرى مجتمعةً وهكذا إلى $r_{1.23\ldots k}$ التي تعني الإرتباط المتعدد بين x_1 مع k من المتغيرات الأخرى مجتمعةً.

يمكن حساب الإرتباط المتعدد بين x_1 من جهة و x_2 فضلاً عن x_3 من جهة أخرى من خلال الصيغة الآتية:

$$r_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12} \cdot r_{13} \cdot r_{23}}{1 - r_{23}^2}} \quad ; \quad r_{23} \neq \mp 1 \quad \cdots \quad (6.9)$$

ويمكن حسابه أيضاً بدلالة الإرتباطات الجزئية وكما يأتي:

$$r_{1.23} = \sqrt{1 - \left(1 - r_{12}^2\right)\left(1 - r_{13.2}^2\right)} \qquad \cdots \quad (6.10)$$

ويمكن حساب $r_{1.234}$ من الإرتباطات الجزئية من خلال الصيغة الآتية:

$$r_{1.234} = \sqrt{1 - \left(1 - r_{12}^2\right)\left(1 - r_{13.2}^2\right)\left(1 - r_{14.23}^2\right)} \qquad \cdots \quad (6.11)$$

وهكذا حتى إلى $r_{1.23\ldots k}$ الذي يمكن أن نحصل عليه من خلال الصيغة الآتية:

$$r_{1.234} = \sqrt{1 - \left(1 - r_{12}^2\right)\left(1 - r_{13.2}^2\right)\left(1 - r_{14.23}^2\right)\cdots\left(1 - r_{1k.23\ldots k-1}^2\right)} \cdots \quad (6.12)$$

مثال (7.6): بالإعتماد على المثال (5.6)، جد معامل الإرتباط المتعدد بين x_1 من جهة و x_2 و x_3 من جهة أخرى. كذلك جد معامل الإرتباط المتعدد بين x_2 من جهة و x_1 و x_3 من جهة أخرى.

الحل: من خلال المثال (5.6) لدينا مايلي:

$r^2_{23.1} = 0.0033$	$r_{13.2} = 0.8257$	$r_{23} = 0.939$	$r_{13} = 0.981$	$r_{12} = 0.957$

الإرتباط المتعدد بين الإنفاق الشهري من جهة والدخل الشهري وعدد أفراد الأسرة من جهة أخرى يمكن أن نحصل عليه من خلال ما يلي:

$$r_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2\, r_{12} \cdot r_{13} \cdot r_{23}}{1 - r_{23}^2}}$$

$$= \sqrt{\frac{0.957^{\;2} + 0.981^{\;2} - 2(0.957)(0.981)(0.939)}{1 - 0.939^{\;2}}}$$

$$= \sqrt{\frac{0.1151}{0.1183}} = 0.9865$$

وهذا يعني أن هنالك إرتباط طردي قوي يساوي 98.65% بين الإنفاق الشهري للأسرة من جهة والدخل الشهري للأسرة وعدد أفراد الأسرة من جهة أخرى.

ويمكن حسابه من خلال الإرتباطات الجزئية وكما يأتي:

$$r_{1.23} = \sqrt{1 - (1 - r_{12}^2)(1 - r_{13.2}^2)}$$

$$= \sqrt{1 - (1 - 0.957^{\;2})(1 - 0.8257^{\;2})}$$

$$= \sqrt{1 - (0.0842)(0.3182)} = 0.9865$$

ويمكن حساب $r_{2.13}$ من خلال الصيغة الآتية:

$$r_{2.13} = \sqrt{1-\left(1-r_{12}^2\right)\left(1-r_{23.1}^2\right)}$$
$$= \sqrt{1-\left(1-0.957^2\right)\left(1-0.0033^2\right)}$$
$$= \sqrt{1-\left(0.0842\right)\left(0.99999\right)} = 0.957$$

وهذا يعني أن هنالك إرتباط طردي قوي يساوي 95.7% بين الدخل الشهري للأسرة من جهة والإنفاق الشهري للأسرة وعدد أفراد الأسرة من جهة أخرى.

مثال (8.6)**: فيمايلي بيانات حول إنتاجية الدونم الواحد من الحنطة x_1 وكمية البذور المستخدمة x_2 وكميات الأمطار الموسمية x_3 ونسبة الملوحة في التربة x_4 :**

x_4	x_3	x_2	x_1
4	**82**	**67**	**23**
3	**98**	**89**	**34**
2	**108**	**99**	**42**
5	**88**	**83**	**35**
1	**110**	**100**	**60**
2	**102**	**90**	**45**

والمطلوب ما يلي:

1- **حساب الإرتباط الخطي البسيط بين كمية إنتاجية الدونم الواحد وكمية البذور المستخدمة.**
2- **حساب الإرتباط الجزئي بين كمية إنتاجية الدونم الواحد وكمية البذور المستخدمة باستبعاد أثر كميات الأمطار الموسمية.**
3- **حساب الإرتباط الجزئي بين كمية إنتاجية الدونم الواحد وكمية البذور المستخدمة باستبعاد أثر كميات الأمطار الموسمية ونسبة الملوحة في التربة.**

4- حساب الإرتباط المتعدد بين كمية إنتاجية الدونم الواحد من جهة وكمية البذور المستخدمة وكميات الأمطار الموسمية ونسبة الملوحة في التربة من جهة أخرى.

الحل:

1- حساب r_{12} هو كما يأتي:

$$r_{12} = \frac{s_{x_1 x_2}}{s_{x_1} s_{x_2}} = \frac{129.6184}{(12.4807)(12.1326)} = 0.856$$

وهذا يعني أن هنالك إرتباط طردي قوي يساوي 85.6% بين كمية إنتاجية الدونم الواحد وكمية البذور المستخدمة.

2- حساب $r_{12.3}$ هو كما يأتي:

نحتاج إلى حساب r_{13} و r_{23} أولاً:

$$r_{13} = \frac{s_{x_1 x_3}}{s_{x_1} s_{x_3}} = \frac{120.3832}{(12.4807)(11.0996)} = 0.869$$

$$r_{23} = \frac{s_{x_2 x_3}}{s_{x_2} s_{x_3}} = \frac{129.5497}{(12.1326)(11.0996)} = 0.962$$

لذلك فإن:

$$r_{12.3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}} = \frac{0.856 - (0.869)\cdot(0.962)}{\sqrt{(1 - 0.869^2)(1 - 0.962^2)}}$$

$$= \frac{0.02}{\sqrt{(0.2448)(0.0746)}} = \frac{0.02}{0.1351} = 0.148$$

وهذا يعني أن هنالك إرتباط طردي ضعيف ويساوي 14.8% بين إنتاجية الدونم الواحد من الحنطة وكمية البذور المستخدمة مع إستبعاد أثر كميات الأمطار الموسمية.

3- حساب $r_{12.34}$ هو كما يأتي:

نحتاج إلى حساب $r_{14.3}$ و $r_{24.3}$ أولاً:

$$r_{14.3} = \frac{r_{14} - r_{13}\, r_{43}}{\sqrt{\left(1 - r_{13}^2\right)\left(1 - r_{43}^2\right)}} = -0.096$$

$$r_{24.3} = \frac{r_{24} - r_{23}\, r_{43}}{\sqrt{\left(1 - r_{23}^2\right)\left(1 - r_{43}^2\right)}} = 0.958$$

لذلك فإن:

$$r_{12.34} = \frac{r_{12.3} - r_{14.3}\, r_{24.3}}{\sqrt{\left(1 - r_{14.3}^2\right)\left(1 - r_{24.3}^2\right)}}$$

$$= \frac{0.148 - (-0.096)(0.958)}{\sqrt{\left(1 - (-0.096)^2\right)\left(1 - 0.958^2\right)}}$$

$$= \frac{0.24}{\sqrt{(0.9908)(0.0822)}} = \frac{0.24}{0.2854} = 0.841$$

وهذا يعني أن هنالك إرتباط طردي قوي ويساوي 84.1% بين كمية إنتاجية الدونم الواحد وكمية البذور المستخدمة بإستبعاد أثر المتغيرين كمية الأمطار الموسمية ونسبة ملوحة التربة.

4- حساب $r_{1.234}$ هو كما يأتي:

نحتاج إلى حساب $r_{13.2}$ و $r_{14.23}$ أولاً:

$$r_{1.234} = \sqrt{1-\left(1-r_{12}^2\right)\left(1-r_{13.2}^2\right)\left(1-r_{14.23}^2\right)}$$
$$= \sqrt{1-\left(1-0.856^2\right)\left(1-0.323^2\right)\left(1-(-0.826)^2\right)}$$
$$= \sqrt{1-(0.2673)(0.8957)(0.3177)} = \sqrt{0.9239} = 0.961$$

وهذا يعني أن هنالك إرتباط طردي قوي جداً يساوي 96.1% بين كمية إنتاجية الدونم الواحد من الحنطة من جهة وكمية البذور المستخدمة وكمية الأمطار الموسمية ونسبة ملوحة التربة مجتمعة من جهة أخرى.

6.6: الإرتباط بين الصفات Correlation between Attributes:

وهو الإرتباط الذي يقترن بوجود توزيع تكراري مزدوج لمتغيرين من النوع الوصفي (كلاهما أو أحدهما من النوع الوصفي) مما يتعذر استخدام معامل إرتباط سبيرمان (لإنها تكون غير قابلة للترتيب تصاعدياً أو تنازلياً) ويستوجب مقياس آخر ملائم لهذه الحالة مثل معامل التوافق أو معامل الإقتران (وهنالك أنواع أخرى لامجال لذكرها هنا) وهي كما يلي:

أ- معامل التوافق Coefficient of Contingency:

يمكن إيجاد معامل التوافق للبيانات المبوبة الوصفية غير القابلة للترتيب (أو قابلة للترتيب) من خلال تكوين جدول التوافق الآتي:

المستويات		**مستويات y**				
		y_1	y_2	$\cdots$	y_m	المجموع
مستويات x	x_1	f_{11}	f_{12}	$\cdots$	f_{1m}	$T_{1.}$
	x_2	f_{21}	f_{22}	$\cdots$	f_{2m}	$T_{2.}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	x_k	f_{k1}	f_{k2}	$\cdots$	f_{km}	$T_{k.}$
المجموع		$T_{.1}$	$T_{.2}$	$\cdots$	$T_{.m}$	$T_{km} = n$

فإن معامل التوافق C يمكن أن نحصل عليه بالإعتماد على صفوف (أو أعمدة) جدول التوافق كـما في الصيغة الآتية:

$$C = \sqrt{\frac{r-1}{r}} \qquad \cdots \quad (6.13)$$

حيث أن r نحصل عليها من خلال جمع قيم الصفوف r_j ، أي أن: $r = \sum_{j=1}^{k} r_j$

وأن قيم r_j نحصل عليها من خلال الصيغة الآتية:

$$r_j = \frac{1}{T_{j.}} \sum_{i=1}^{m} \frac{f_{ji}^{2}}{T_{.i}} \qquad \cdots \quad (6.14)$$

مثال (9.6)**: الجدول التالي يبيّن عدد حوادث الطرق التي وقعت في إحدى المدن خلال فترة زمنيـة موزعـة حسب نوع الحادث وحالة الطقس، والمطلوب حساب معامل التوافق.**

حالة الطقس	نوع الحادث	دهس	اصطدام	إنقلاب	المجموع
صحو		15	8	5	28
ممطر		5	25	15	45
ضباب		10	23	20	53
المجموع		30	56	40	126

الحل: نحسب أولاً قيم r_j من الصيغة (6.14) وكما يأتي:

$$r_j = \frac{1}{T_{j.}} \sum_{i=1}^{m} \frac{f_{ji}^2}{T_{.i}}$$

$$r_1 = \frac{1}{T_{1.}} \sum_{i=1}^{3} \frac{f_{1i}^2}{T_{.i}} = \frac{1}{28}\left(\frac{15^2}{30} + \frac{8^2}{56} + \frac{5^2}{40}\right) = \frac{9.2679}{28} = 0.331$$

$$r_2 = \frac{1}{T_{2.}} \sum_{i=1}^{3} \frac{f_{2i}^2}{T_{.i}} = \frac{1}{45}\left(\frac{5^2}{30} + \frac{25^2}{56} + \frac{15^2}{40}\right) = \frac{17.619}{45} = 0.3915$$

$$r_3 = \frac{1}{T_{3.}} \sum_{i=1}^{3} \frac{f_{3i}^2}{T_{.i}} = \frac{1}{53}\left(\frac{10^2}{30} + \frac{23^2}{56} + \frac{20^2}{40}\right) = \frac{22.7798}{53} = 0.4298$$

لذلك فإن:

$$r = \sum_{j=1}^{k} r_j = r_1 + r_2 + r_3 = 0.331 + 0.3915 + 0.4298 = 1.1523$$

فإن معامل التوافق نحصل عليه من خلال (6.13) وكما يأتي:

$$C = \sqrt{\frac{r-1}{r}} = \sqrt{\frac{1.1523-1}{1.1523}} = 0.3636$$

وهذا يعني أن هنالك نسبة توافق مقدارها 36.36% بين حالة الطقس ونوع الحادث.

ملاحظة: إن قيمة معامل التوافق تتراوح بين الصفر والواحد ولايمكن حسابه عندما تتراوح قيمة r مابين الصفر والواحد.

ب- معامل الإقتران Coefficient of Association**:**

هنالك حالتان في عملية حسابه وهي كما يلي:

أولاً: في حالة وجود مستويين لكل متغير:

معامل الإقتران يقيس العلاقة بين متغيرين وصفيين غير قابلين للترتيب (أو ربما أحدهما أوكلاهما قابلين للترتيب) مفرغة بياناتهما في جدول توافق بحجم 2 × 2 مثل بين الجنس (ذكر، أنثى) وكفاءة الأداء في العمل (جيد، سيىء) وجدول التوافق يكون كالآتي:

المستويات		مستويات y		
مستويات x		y_1	y_2	المجموع
	x_1	f_{11}	f_{12}	$f_{1.}$
	x_2	f_{21}	f_{22}	$f_{2.}$
	المجموع	$f_{.1}$	$f_{.2}$	n

وبالإعتماد على جدول التوافق نحصل على معامل الإقتران من خلال تطبيق الصيغة الآتية:

$$C.A_2 = \frac{f_{11} \cdot f_{22} - f_{12} \cdot f_{21}}{f_{11} \cdot f_{22} + f_{12} \cdot f_{21}} \qquad \cdots \quad (6.15)$$

مثال (10.6)**: الجدول التالي يبيّن عدد الذكور والإناث المشتركين في إمتحان كفاءة التدريس والمطلوب حساب معامل الإقتران بين الجنس وكفاءة التدريس.**

الجنس	كفاءة التدريس	جيد	سيىء	المجموع
الذكور		40	25	65
الإناث		30	15	45
المجموع		70	40	110

الحل: يمكن حساب معامل الإقتران من خلال الصيغة (6.15) وما يلي:

$$C.A_2 = \frac{f_{11} \cdot f_{22} - f_{12} \cdot f_{21}}{f_{11} \cdot f_{22} + f_{12} \cdot f_{21}}$$

$$= \frac{(40)(15)-(25)(30)}{(40)(15)+(25)(30)} = \frac{-150}{1350} = -0.1111$$

وهذا يعني هنالك إرتباط عكسي ضعيف بين الجنس وكفاءة التدريس.

ملاحظة: قيمة معامل الإقتران تكون محصوره مابين 1 و -1

ثانياًً: في حالة وجود مستويين أو أكثر لكل متغير:

يمكن حساب معامل الإقتران بالإعتماد على جدول التوافق k × m الموضح آنفاً ومن خلال تطبيق الصيغة الآتية:

$$C.A_G = \frac{\sum_{i=1}^{m} f'_{ij} + \sum_{j=1}^{k} f''_{ij} - T'_{.i} - T'_{j.}}{2n - \left(T'_{.i} + T'_{j.}\right)} \qquad \cdots \quad (6.16)$$

حيث أن:

f'_{ij} يمثل أكبر تكرار في العمود i

f''_{ij} يمثل أكبر تكرار في الصف j

$T'_{.i}$ يمثل أكبر مجموع من بين مجاميع الأعمدة.

$T'_{j.}$ يمثل أكبر مجموع من بين مجاميع الصفوف.

مثال (11.6)**: لبيانات جدول التوافق في المثال** (9.6) **أحسب معامل الإقتران.**

الحل: يتم تكوين الجدول الآتي:

حالة الطقس / نوع الحادث	دهس	اصطدام	إنقلاب	المجموع	أكبر تكرار في العمود	أكبر مجموع للصفوف
صحو	15	8	5	28	15	---
ممطر	5	25	15	45	25	---
ضباب	10	23	20	53	23	53
المجموع	30	56	40	126	63	---
أكبر تكرار في العمود	15	25	20	60		
أكبر مجموع للأعمدة	---	56	---	---		

من خلال جدول التوافق وتطبيق الصيغة (6.16) نحصل على معامل الإقتران وكما يأتي:

$$C.A_G = \frac{\sum_{i=1}^{m} f'_{ij} + \sum_{j=1}^{k} f''_{ij} - T'_{.i} - T'_{j.}}{2n - \left(T'_{.i} + T'_{j.}\right)}$$

$$= \frac{60 + 63 - 56 - 53}{2(126) - (56 + 53)} = \frac{14}{143} = 0.0979$$

وهنالك أنواع أخرى من معاملات التوافق والإقتران ذات ثلاثة إتجاهات أو أربعة وغيرها لامجال لذكرها في هذا الكتاب.

تمارين الفصل السادس

6-1: ما المقصود بمعامل الإرتباط الخطي البسيط ؟

6-2: جد معامل الإرتباط الخطي البسيط بين الإيرادات والنفقات خلال فترة معينة لدائرة كهرباء أربيل المبينة من خلال الجدول الآتي:

إيرادات	23	34	46	33	50	55	66
نفقات	12	25	30	22	35	40	45

6-3: جد معامل الإرتباط الخطي البسيط بين كفاءة الممرضات من خلال إمتحان كفاءة خصص لسبعة ممرضات مقابل عدد سنوات الخبرة والتي كانت كما يأتي:

الكفاءة x	ممتاز	جيد	جيد جداً	ضعيف	جيد جداً	متوسط	مقبول
سنوات الخبرة y	20	15	18	7	12	12	12

6-4: جد معامل الإرتباط الخطي البسيط بين تقديرات إمتحان كفاءة اللغة الانكليزية وتحصيلهم الدراسي لستة متقدمين لوظيفة ما وكانت كما يأتي:

الكفاءة x	متوسط	جيد جداً	ضعيف	جيد	ممتاز	مقبول
التحصيل الدراسي y	إبتدائية	متوسطة	إعدادية	ماجستير	دكتوراه	بكالوريوس

6-5: الجدول الآتي يبين البيانات لخمسة من المديرين تتضمن، المسمى الوظيفي A، المؤهل العلمي B، الخبرة العملية C:

المسمى الوظيفي A	1	2	3	4	5
المؤهل العلمي B	3	4	2	1	5
الخبرة العملية C	2	4	9	13	12

والمطلوب ما يلي:

1- **جد معامل الإرتباط الجزئي بين المسمى الوظيفي والمؤهل العلمي بإستبعاد أثر الخبرة العملية.**

2- **جد معامل الإرتباط الجزئي بين المؤهل العلمي والخبرة العملية بإستبعاد أثر المسمى الوظيفي.**

3- **جد معامل الإرتباط الجزئي بين المسمى الوظيفي والخبرة العملية بإستبعاد أثر المؤهل العلمي.**

6-6: إذا توفرت لديك البيانات الآتية:

x	2	3	4	5	6
y	6	4	2	0	2-
z	20	34	42	65	78
v	0	3-	2	5-	1

جد ما يلي:

1	2	3
$r_{xy.zv}$	$r_{xz.yv}$	$r_{vz.xy}$

6-7: الآتي نتائج إختبار نوعين من البذور وتأثيرها في زيادة كمية محصول الطماطة بفرض ثبات الظروف التجريبية الأخرى في 200 قطعة زراعية.

كمية المحصول	نوع البذور	محسنة	إعتيادية
فوق المتوسط		70	40
دون المتوسط		55	35

والمطلوب حساب معامل التوافق والإقتران بين كمية المحصول ونوع البذور.

6-8: الآتي توزيع تكراري مزدوج لعدد من الزلازل التي وقعت في أنحاء متفرقة من العالم خلال فترة زمنية معينة موزعة حسب قوة الزلزال على مقياس ريختر (المؤلف من تسع درجات) والأضرار الناجمة عن الزلزال.

الأضرار	الدرجة	أقل من 5	5-	6-	7-	8-9
طفيفة		12	4	2	--	--
متوسطة		--	1	4	7	--
جسيمة		--	--	1	10	13

جد معامل التوافق والإقتران بين قوة الزلزال والأضرار الناجمة عنه.

الفصل السابع

تحليل الإنحدار

Regression Analysis

الفصل السابع

تحليل الإنحدار

Regression Analysis

1.7: مقدمة:

تناولنا سابقاً معامل الإرتباط الذي يقيس مقدار واتجاه الترابط بين متغيرين ولكن ربما احتاج الباحث الى معرفة أكثر من مجرد العلاقة الإرتباطية بين هذين المتغيرين فمثلاً ربما احتاج الى توقع تنبؤ (Prediction) سلوك أحد المتغيرات في ضوء تأثره بمتغير آخر أو بعدة متغيرات أخرى. كذلك فإنه اذا كان هناك عدة متغيرات فمن المحتمل أن تكون لدى الباحث رغبة في تقرير مدى تأثير كل متغير من هذه المتغيرات على متغير آخر، إن هذا المتغير الذي يرغب الباحث في دراسة سلوكه ومعرفة مدى تأثره بالمتغيرات الأخرى يسمى بالمتغير التابع أو المعتمد (Dependent variable) ويطلق على المتغير الآخر الذي يؤثر في سلوك المتغير التابع بالمتغير المستقل (Independent variable) لذلك فإن تحليل الإنحدار يعتمد على وجود متغير تابع (معتمد) ومتغير أو مجموعة من المتغيرات المستقلة.

2.7: أهداف تحليل الانحدار:

1- دراسة العلاقة بين متغيرين على شكل علاقة دالية بحيث يمكن معرفة التغير في أحدهما على أساس تأثره بالآخر، أي أن الدراسة تكون لهدف توقع وتنبؤ سلوك المتغير المعتمد في ضوء تأثره بالمتغير أو المتغيرات المستقلة، أي أنه يمكن تنبؤ التغيرات الحاصلة في المتغير التابع على أساس معرفة التغيرات الحاصلة في المتغيرات المستقلة أو المتنبأة، وتكتب العلاقة الدالية $y = f(x)$.

تقرير مدى مساهمة كل متغير مستقل في مدى التباين أو التغيرات الحاصلة في المتغير المعتمد.

قياس مدى الترابط الكلي بين المتغير المعتمد والمتغيرات المستقلة.

إجراء سلسلة من الإختبارات الفرضية لأي من العلاقات المشار اليها في النقاط السابقة.

ملاحظة: لابد من التفريق بين تحليل الإرتباط وتحليل الانحدار حتى لايحدث خلط بينهما، فعلى الرغم من تشابه العلاقة الرياضية بين الإرتباط والانحدار الا أنهما يختلفان عن بعضهما في النواحي التالية:

يشترط تحليل الانحدار أن تكون هناك علاقة دالية بين المتغيرات بينما لايوجد مثل هذا الشرط في تحليل الإرتباط.

يشترط في حالة الإرتباط أن تكون بيانات كل المتغيرات موزعة توزيعا طبيعياً بينما يشترط تحليل الانحدار أن تتوزع قيم المتغير المعتمد y بشكل طبيعي.

ان تحليل الإرتباط هو عبارة عن مقياس وصفي بينما تحليل الانحدار هو مقياس كمي.

هنالك أمثلة عديدة يمكن تطبيقها على موضوع تحليل الإنحدار مثل علاقة إرتباطية بين جودة الانتاج الزراعي وكميات الأمطار الساقطة، أي أنه كلما زادت كميات الأمطار توقعنا زيادة في الانتاج الزراعي، كذلك دراسة مدى تأثير التغيرات الحاصلة في الدخل الشهري للعائلة على التغيرات الحاصلة في الإنفاق الشهري للعائلة، ودراسة تأثير السكائر على أمراض الرئة، كذلك دراسة العلاقة بين سرعة الإنجاز في السباحة ومساحة كف اللاعب. و... الخ.

هنالك أنواع عديدة من الإنحدار منها البسيط (متغير مستقل واحد) ومنها المتعدد (أكثر من متغير مستقل واحد) ومنها الخطي (المتغير المستقل ذات أس مساوي للواحد) ومنها غير الخطي (المتغير المستقل ذات أس أكبر من واحد أو تظهر بهيئة دالة لوغارتمية أو غيرها).

3.7: الانحدار الخطي البسيط Simple Linear Regression:

الإنحدار الخطي البسيط هو عبارة عـن عمليـة تقـدير العلاقـة الخطيـة بـين متغـيرين أحـدهما مستقل والآخر معتمد. كما في نموذج الإنحدار الآتي:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad \cdots \quad (7.1)$$

يمثل y_i مشاهدات المتغير المعتمد (مثل الإنفاق الشهري للعائلة) في حين يمثل x_i مشاهدات المتغير المستقل (مثل الدخل الشهري للأسرة)، في حين يمثل ε_i الخطأ العشوائي (يجب وجوده في النموذج لأنه ليس لدينا علاقة تامة بين المتغيرين) والذي يمثل التغيرات الحاصلة في المتغير المعتمد نتيجة تـأثيرات عدة متغـيرات أخـرى (عـدى المتغـير المـستقل) غـير مدروسـة. وأن β_0 و β_1 تمثـل معلـمات نمـوذج الإنحدار الخطي البسيط للمجتمع وتـستخدم أحـدى طرائـق التقـدير في تقـدير معلـمات خـط الانحـدار البسيط للعينة مثل طريقة المربعات الصغرى الاعتيادية (OLS)، أي تقدير معلمات النموذج الآتي:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \qquad \cdots \quad (7.2)$$

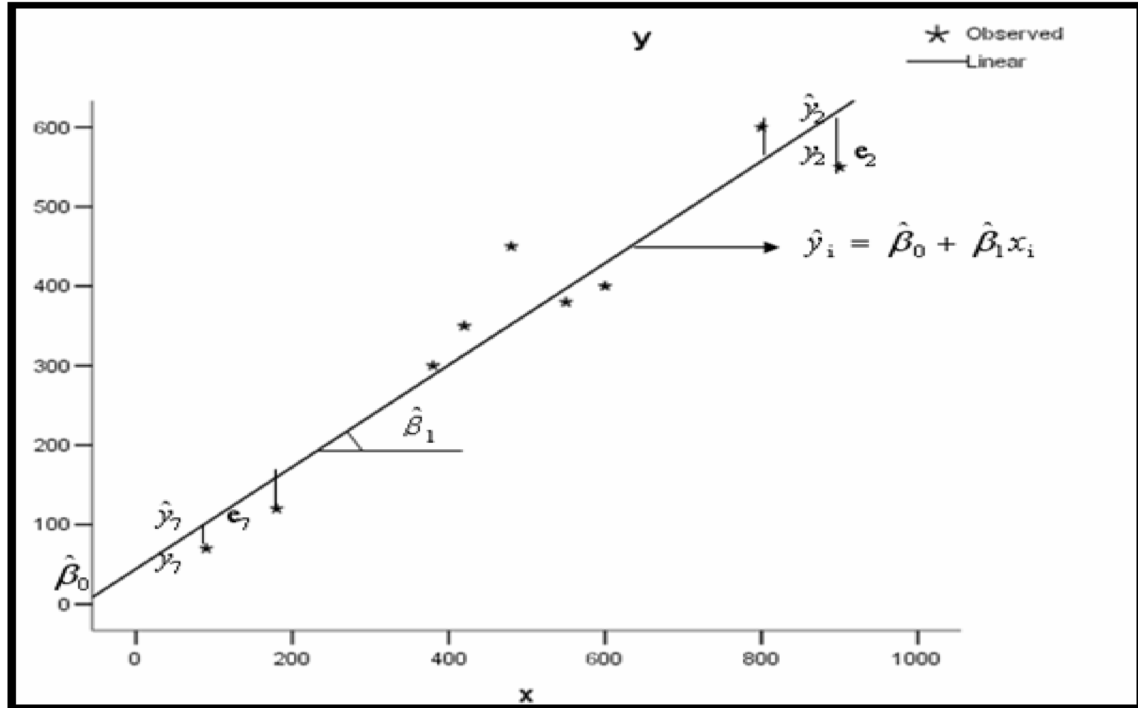
$\hat{y}_i$ تمثل القيم المقدرة للمتغير المعتمد من خلال النموذج المقدر أعـلاه وأن $\hat{\beta}_0$ و $\hat{\beta}_1$ تمثـل معلمات النموذج المقدرة من العينة والتي يمكن حسابها من خلال الصيغتان الآتيتين:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_x^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\cdot\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2} \qquad \cdots \quad (7.3)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \qquad \cdots \quad (7.4)$$

هاتين القيمتين تحددان أفضل خط مستقيم يتخلل الشكل الإنتشاري معبراً عن العلاقة مـابين x و y. وأن $\hat{\beta}_0$ تمثل نقطة تقاطع خط الإنحدار مع المحور العمودي (الذي يمثل المتغير y) أو تمثل المسافة مابين نقطة الأصل (0,0) ونقطة تقاطع خط الإنحدار مع المحور العمودي. في حين تمثـل $\hat{\beta}_1$ ميـل (Slop) الإنحدار أو معامل الإنحدار Regression Coefficient أي مقياس يوضح مقدار تغير y إذا ماتغيرت x بوحـدة واحدة، وأن قيمة هذه المعلمة قد تكون موجبة أو سالبة تبعاً لقيمة التباين المشترك (قيمـة البـسط) فـإذا كانت موجبة دلَ ذلك على وجود علاقة طردية بين المتغـيرين أمـا إذا كانـت قيمتهـا سـالبة دلَ ذلـك عـلى وجود علاقة عكسية بين المتغيرين في حين إذا كانت قيمتها مساوية للصفر دلَ ذلك على عدم وجود علاقـة بين المتغيرين. والشكل التالي يوضح العلاقة الخطية والشكل الإنتشاري بين الدخل الشهري للعائلة والإنفاق الشهري للعائلة لعينة عشوائية تتألف من تسعة عوائل كما في الجدول الآتي:

	x	y
1	600	400
2	900	550
3	550	380
4	180	120
5	420	350
6	380	300
7	90	70
8	800	600
9	480	450



الشكل (7.1)

يبين الشكل الإنتشاري وخط الإنحدار للعلاقة بين الدخل والإنفاق الشهري

حيث تمثل قيم y_i المشاهدات الحقيقية للعينة في حين تمثـل $\hat{y}_i$ القـيم المقـدرة لهـا باسـتخدام نموذج الإنحدار الخطي البسيط والفرق بيـنهما يـدعى بـالبواقي Residuals وهـو عبـارة عـن الفـرق مـابين القيمة الحقيقية إلى y_i والمقدرة لها $\hat{y}_i$ أي أن:

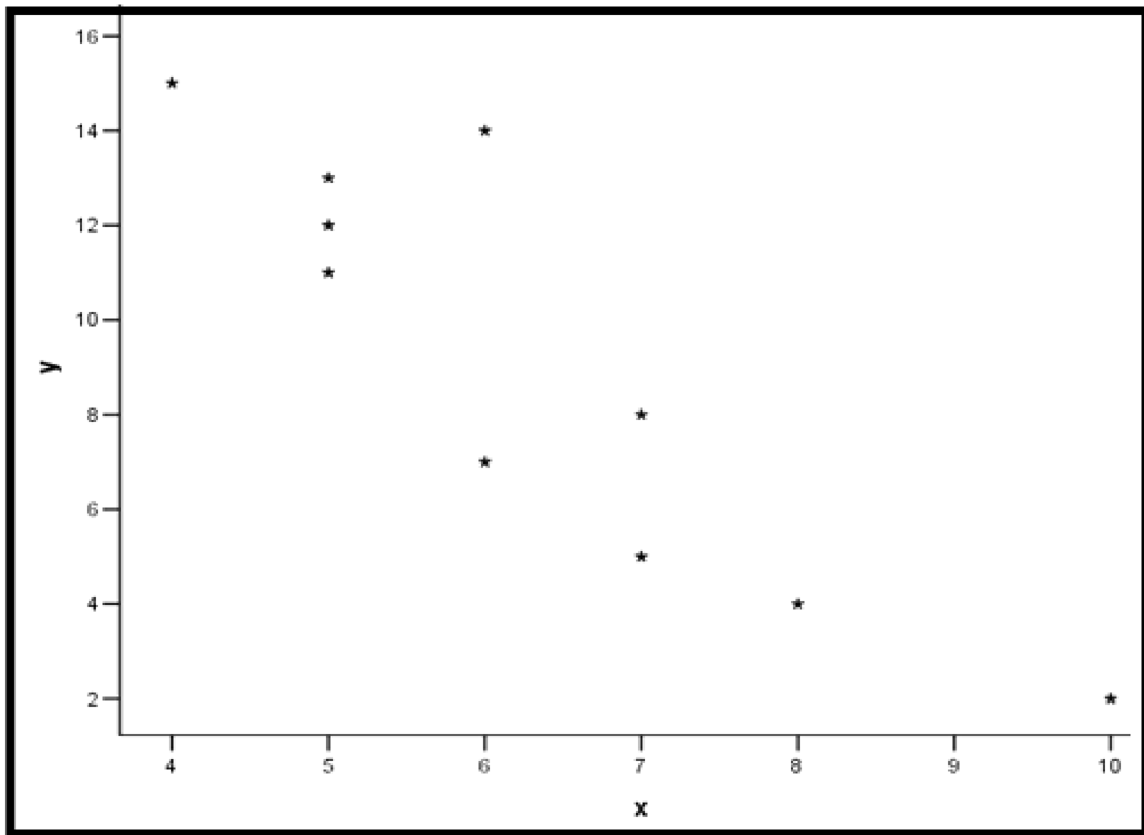
$$e_i = y_i - \hat{y}_i \qquad \cdots \quad (7.5)$$

إن طريقة المربعـات الـصغرى المـستخدمة في تقـدير معلـمات الإنحـدار الخطـي البـسيط تجعـل مجموع مربعات البواقي أصغر مايمكن أي جعل $\sum_{i=1}^{n} e_i^2$ أقل مايمكن.

مثال (1.7)**: البيانات التالية تمثل الكمية المطلوبة y_i من سلعة معينة وسعر الوحدة الواحدة منها x_i، يطلب رسم الإنتشار وتقدير نموذج الإنحدار الخطي البسيط y/x.**

y_i	2	4	7	5	12	15	13	14	11	8
x_i	10	8	6	7	5	4	5	6	5	7

الحل: الرسم إنتشار البيانات يجب تحديد أزواج القيم (x_i , y_i) بالشكل (2,10)، (4,8)... (8,7) ويمثل x_i المحور الأفقي في حين يمثل y_i المحور العمودي فنحصل على الشكل الآتي:



الشكل (7.2)

يبين الشكل الإنتشاري بين الكمية المطلوبة من السلعة وسعرها

مــن خــلال الــشكل أعــلاه نلاحــظ أن هنالــك علاقــة خطيــة عكــسية بــين الكميــة

المطلوبة من السلعة وسعر الوحدة الواحدة منها.

ولتقدير النموذج الخطي يجب أولاً حساب الأوساط الحسابية للمتغيرين وكما يأتي:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{10} y_i}{10} = \frac{91}{10} = 9.1$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{63}{10} = 6.3$$

ولحساب معلمات النموذج يجب تكوين الجدول الآتي:

التسلسل	y_i	x_i	$y_i - \bar{y}$	$x_i - \bar{x}$	$(y_i - \bar{y})(x_i - \bar{x})$	$(x_i - \bar{x})^2$
1	2	10	7.1-	3.7	26.27-	13.69
2	4	8	5.1-	1.7	8.67-	2.89
3	7	6	2.1-	0.3-	0.63	0.09
4	5	7	4.1-	0.7	2.87-	0.49
5	12	5	2.9	1.3-	3.77-	1.69
6	15	4	5.9	2.3-	13.57-	5.29
7	13	5	3.9	1.3-	5.07-	1.69
8	14	6	4.9	0.3-	1.47-	0.09
9	11	5	1.9	1.3-	2.47-	1.69
10	8	7	1.1-	0.7	0.77-	0.49
المجموع			**0**	**0**	**64.3 -**	**28.10**

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{-64.3}{28.10} = -2.2883$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 9.1 - (-2.2883)(6.3) = 23.5163$$

وهذا يعني أن النموذج المقدر هو كالآتي:

$$\hat{y}_i = 23.5163 - 2.2883\, x_i$$

من خلال النموذج أعلاه وبالتحديد قيمة $\hat{\beta}_1 = -2.2883$ نلاحظ أنها قيمة سالبة مما تدل على انه هنالك علاقة عكسية بين الكمية المطلوبة من السلعة وسعر الوحدة الواحدة منها. وأن زيادة سعر هذه السلعة بمقدار وحدة واحدة سيؤدي إلى إنخفاض الكمية المطلوبة بمقدار 2.2883، وإحدى استخدامات النموذج المقدر هو التنبؤ بقيمة y عندما تكون x ذات قيمة معينة فإذا إفترضنا أن قيمة سعر السلعة هو $x_0 = 3$ يمكن من خلال النموذج المقدر تقدير الكمية المطلوبة من هذه السلعة وكما يأتي:

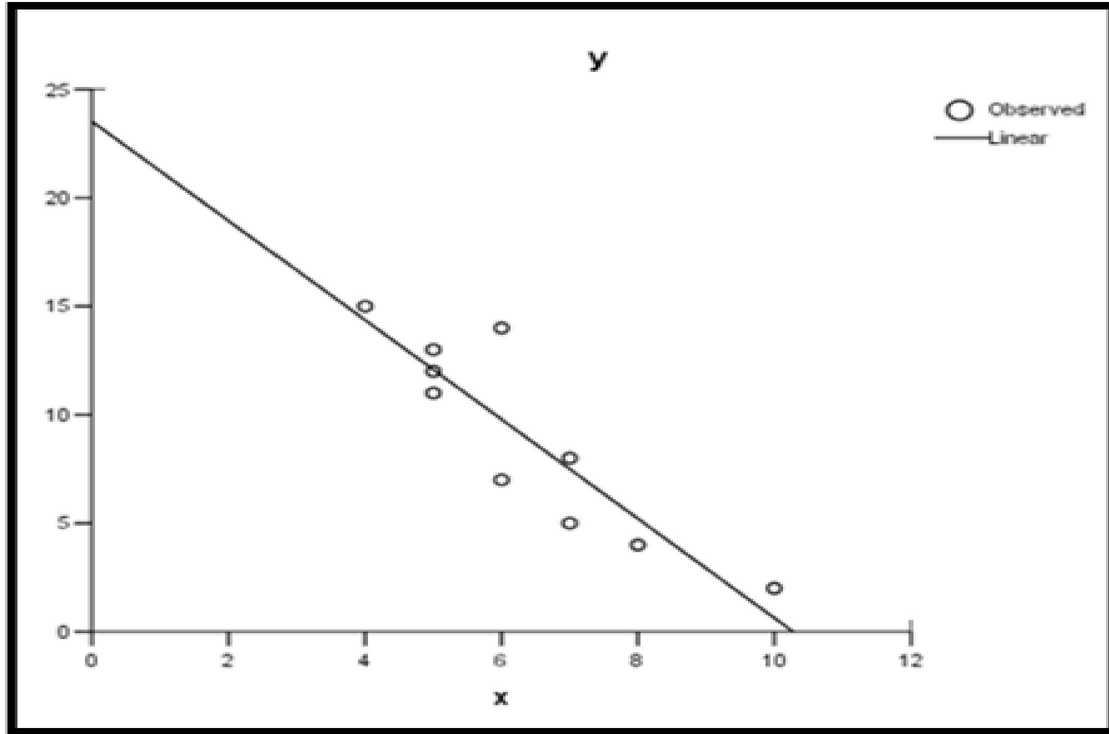
$$\hat{y}_i = 23.5163 - 2.2883x_i$$

$$\Rightarrow \quad \hat{y}_0 = 23.5163 - 2.2883x_0$$

$$= 23.5163 - 2.2883(3) = 16.6514$$

وهذا يعني أنه إذا كان سعر السلعة (3) فإن الكمية المطلوبة من هذه السلعة هو (16.6514)

كما يمكن رسم الإنتشار مع خط الإنحدار للنموذج المقدر الذي يوضح العلاقة بين الكمية المطلوبة وسعر الوحدة الواحدة منها وكما يأتي:



الشكل (7.3)

يبين الشكل الإنتشاري وخط الإنحدار بين الكمية المطلوبة من السلعة وسعرها

ملاحظة: هنالك علاقة بين معامل الإنحدار $\hat{\beta}_1$ ومعامل الإرتباط البسيط وهي كما يأتي:

$$\hat{\beta}_1 = \frac{S_y}{S_x} \cdot r_{xy} \qquad \cdots \qquad (7.6)$$

وهذا يعني أن:

$$r_{xy} = \frac{S_x}{S_y} \cdot \hat{\beta}_1 \qquad \cdots \qquad (7.7)$$

مثال (2.7)**: في عينة مكونة من** (12) **عاملاً تم قياس معدلات أدائهم** y_i **وأن** x_i **تمثل عدد سنوات خبرة العمل السابقة وكما يأتي:**

معدل الأداء y_i	85	74	76	90	85	87	94	98	81	91	76	74
سنوات الخبرة x_i	1	7	5	2	6	3	2	5	4	3	1	4

يطلب ما يأتي:

1- **رسم الإنتشار بين معدل الأداء وسنوات الخبرة.**

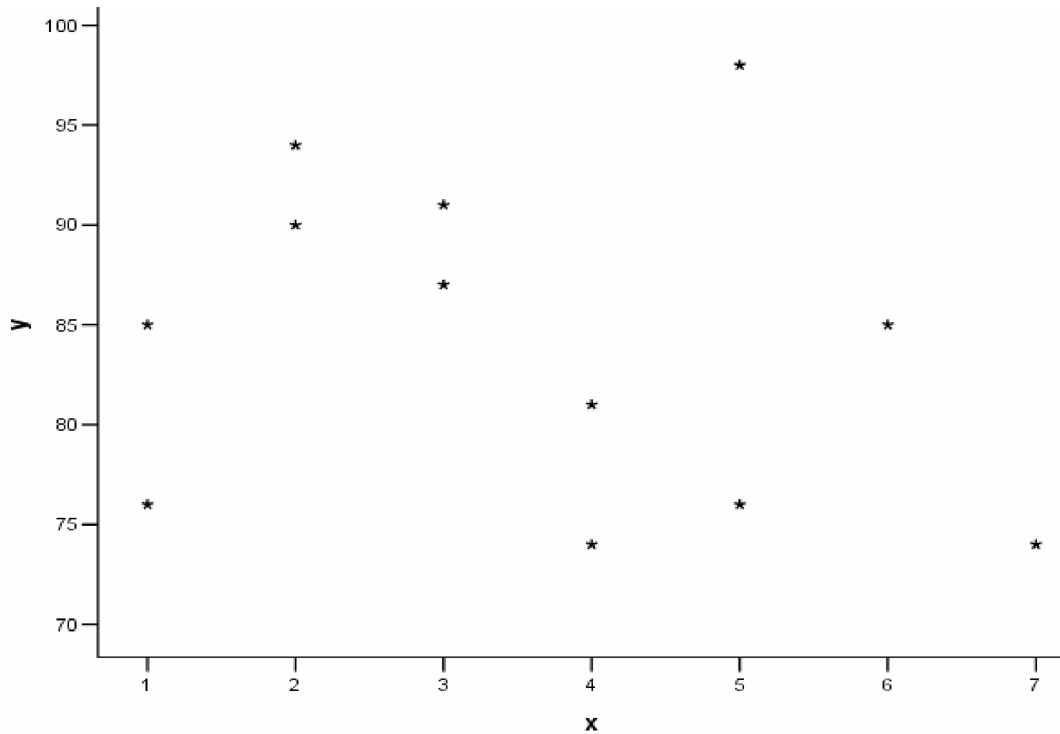
2- **تقدير نموذج الإنحدار الخطي البسيط** y/x.

3- **أثبت حسابياً أن مجموع البواقي يساوي صفر.**

4- **جد معامل الإرتباط الخطي البسيط من معامل الإنحدار.**

الحل:

شكل الإنتشار بين معدل الأداء الذي يمثل المحور العمودي وسنوات الخبرة التي تمثل المحور الأفقي هي كما يأتي:



الشكل (7.4)

يبين الشكل الإنتشاري بين معدل الأداء وسنوات الخبرة

من خلال الشكل الإنتشاري نلاحظ أنه لاتوجد علاقة خطية بين معدل الأداء وسنوات الخبرة بالإعتماد على هذه العينة، لأن توزيع النقاط بشكل عشوائي وليس خطي بإتجاه معين.

2- لتقدير النموذج الخطي بين معدل الأداء وسنوات الخبرة نحسب أولاً الوسط الحسابي لهما:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{12} y_i}{12} = \frac{1011}{12} = 84.25$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{12} x_i}{12} = \frac{43}{12} = 3.5833$$

ثم نكون الجدول الآتي:

التسلسل	y_i	x_i	$x_i y_i$	x_i^2	y_i^2
1	85	1	85	1	7225
2	74	7	518	49	5476
3	76	5	380	25	5776
4	90	2	180	4	8100
5	85	6	510	36	7225
6	87	3	261	9	7569
7	94	2	188	4	8836
8	98	5	490	25	9604
9	81	4	324	16	6561
10	91	3	273	9	8281
11	76	1	76	1	5776
12	74	4	296	16	5476
المجموع	**1011**	**43**	**3581**	**195**	**85905**

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\cdot\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} = \frac{3581-(12)(3.5833)(84.25)}{195-(12)(3.5833)^2} = \frac{-41.7163}{40.9195} = -1.0195$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 84.25-(-1.0195)(3.5833) = 87.9032$$

وهذا يعني أن نموذج الإنحدار الخطي المقدر هو كمايأتي:

$$\hat{y}_i = 87.9032 - 1.0195\, x_i$$

أثبت حسابياً أن مجموع البواقي يساوي صفر.

لحساب البواقي يجب أولاً إيجاد قيم $\hat{y}_i$ المقدرة من النموذج أعلاه وكما يأتي:

$$\hat{y}_i = 87.9032 - 1.0195 x_i$$
$$\hat{y}_1 = 87.9032 - 1.0195 x_1 = 87.9032 - 1.0195(1) = 86.8837$$
$$\hat{y}_2 = 87.9032 - 1.0195 x_2 = 87.9032 - 1.0195(7) = 80.7667$$
$$\vdots$$
$$\hat{y}_{12} = 87.9032 - 1.0195 x_{12} = 87.9032 - 1.0195(4) = 83.8252$$

ولحساب قيم البواقي نستخدم الصيغة (7.5):

$$e_i = y_i - \hat{y}_i$$
$$e_1 = y_1 - \hat{y}_1 = 85 - 86.8837 = -1.8837$$
$$e_2 = y_2 - \hat{y}_2 = 74 - 80.7667 = -6.7667$$
$$\vdots$$
$$e_{12} = y_{12} - \hat{y}_{12} = 74 - 83.8285 = -9.8285$$

وعلى هذا الأساس فإن مجموع البواقي يساوي صفر، أي أن:

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{12} e_i = e_1 + e_2 + \cdots + e_{12}$$
$$= -1.8837 - 6.7667 + \cdots - 9.8285 = 0$$

لحساب معامل الإرتباط الخطي البسيط من معامل الإنحدار نستخدم الصيغة (7.7) والتي تحتاج إلى حساب الإنحراف المعياري لكلا المتغيرين وكما يأتي:

$$S_y=\sqrt{\frac{\sum_{i=1}^{n}y_i^2-n\bar{y}^2}{n-1}}=\sqrt{\frac{85905-(12)(84.25)^2}{12-1}}=\sqrt{\frac{728.25}{11}}=8.1366$$

$$S_x=\sqrt{\frac{\sum_{i=1}^{n}x_i^2-n\bar{x}^2}{n-1}}=\sqrt{\frac{195-(12)(3.5833)^2}{12-1}}=\sqrt{\frac{40.9195}{11}}=1.9287$$

لذلك فإن معامل الإرتباط الخطي البسيط هو كمايأتي:

$$r_{xy}=\frac{S_x}{S_y}\cdot\hat{\beta}_1=\frac{1.9287}{8.1366}\cdot(-1.0195)=-0.2417$$

وهذا يعني أن وفق العينة أعلاه فإن هنالك إرتباط خطي عكسي ضعيف يساوي -24.17% بين معدل الأداء وعدد سنوات الخبرة. ولكن كما ذكرنا آنفاً لاتوجد علاقة خطية بين المتغيرين لهذه العينة (غير معنوية) أي ربما تكون هنالك علاقة غير خطية أو هنالك سبب آخر يؤثر على معدل الأداء.

1.3.7: معامل التحديد Coefficient of Determination:

هو عبارة عن مقياس يعبر عن مقدار مايفسره المتغير المستقل في التغيرات الحاصلة في المتغير المعتمد. ويرمز له عادة R^2 ويمكن الحصول عليه من خلال تربيع معامل الإرتباط الخطي البسيط، أي أن:

$$R^2=(r_{xy})^2 \qquad \cdots \quad (7.8)$$

أو يمكن حسابه من خلال الصيغة الآتية:

$$R^2=1-\frac{\sum_{i=1}^{n}e_i^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} \qquad \cdots \quad (7.9)$$

أو يمكن حسابه من خلال الصيغة الآتية:

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad \cdots \quad (7.10)$$

علماًً أن: $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i \quad and \quad \bar{y} = \bar{\hat{y}}$

وبما أن قيمة معامل الإرتباط الخطي البسيط محصورة مابين(1,1-) لذلك فإن قيمة معامل التحديد تكون محصورة مابين(0,1) .

إن R^2 تمثل مقدار مايفسر المتغير المستقل x من التغيرات الحاصلة في المتغير المعتمد y لذلك فإن $(1-R^2)$ يمثل مقياس يعبر عن مقدار ماتفسره مجموعة المتغيرات الأخرى غير المدروسة في التغيرات الحاصلة في المتغير المعتمد y والتي لاتعزى إلى x. وهذا يعني أنه كلما أرتفعت قيمة R^2 دلَ ذلك على أن x ذات تأثير كبير على المتغير y والعكس صحيح أيضاً، وأن R^2 إذا كانت تساوي الواحد فذلك يعني أن x هو المتغير الوحيد الذي يفسر التغيرات الحاصلة في y في حين إذا كانت قيمة R^2 مساويه للصفر دلَ ذلك على أن التغيرات الحاصلة في y سببها متغيرات مستقلة أخرى وأن x لايمتلك أي تأثير على المتغير المعتمد y

مثال (3.7)**: إذا توفرت لديك المعلومات الآتية:**

$n = 20$	$\sum_{i=1}^{n} y_i = 80$	$\sum_{i=1}^{n} x_i = 50$	$S_y^2 = 200$	$S_x^2 = 100$	$S_{xy} = 40$

والمطلوب مايأتي:

1- تقدير نموذج الإنحدار الخطي البسيط.

-2 حساب معامل الإرتباط الخطي البسيط.

-3 حساب معامل التحديد وتفسير معناه.

-4 ماقيمة y إذا كانت قيمة x تساوي 10

الحل:

1- يمكن تقدير نموذج الإنحدار الخطي البسيط كمايأتي:

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{20} y_i}{20} = \frac{80}{20} = 4$$

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{20} x_i}{20} = \frac{50}{20} = 2.5$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_x^2} = \frac{40}{100} = 0.4$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} = 4 - (0.4)(2.5) = 3$$

لذلك فإن نموذج الإنحدار المقدر هو كما يأتي:

$$\hat{y}_i = 3 + 0.4 x_i$$

من خلال النموذج أعلاه نلاحظ أن قيمة معامل الإنحدار موجبة مما يدل على أنه هنالك علاقة طردية بين المتغير المستقل x والمتغير المعتمد y

2- يمكن حساب معامل الإرتباط الخطي البسيط كمايأتي:

$$r_{xy} = \frac{S_x}{S_y} \cdot \hat{\beta}_1 = \sqrt{\frac{S_x^2}{S_y^2}} \cdot \hat{\beta}_1 = \sqrt{\frac{100}{200}} \cdot (0.4) = 0.2828$$

أو حساب معامل الإرتباط الخطي البسيط بطريقة أخرى وهي:

$$r_{xy} = \frac{S_{xy}}{S_x \cdot S_y} = \frac{40}{\sqrt{100} \cdot \sqrt{200}}$$

$$= \frac{40}{(10) \cdot (14.1421)} = \frac{40}{141.421} = 0.2828$$

وهذا يعني أن هنالك علاقة طردية ضعيفة تساوي 28.28% بين المتغيرين.

حساب معامل التحديد هو كما يأتي:

$$R^2 = (r_{xy})^2 = (0.2828)^2 = 0.08$$

وهذا يعني أن المتغير المستقل x يفسر 8% فقط من التغيرات الحاصلة في المتغير المعتمد y والباقي 92% سببها متغيرات أخرى غير مدروسة، وهذا مايؤكد على ضعف العلاقة مابين المتغيرين وفق المعلومات المعطاة.

قيمة y المتوقعة عندما تكون قيمة x تساوي 10 نحصل عليها من خلال النموذج المقدر في المطلوب الأول وكما يأتي:

$$\hat{y}_i = 3 + 0.4 x_i$$

$$\hat{y}_0 = 3 + 0.4 x_0$$

$$= 3 + 0.4(10) = 7$$

وهذا يعني أن قيمة y المتوقعة تساوي 7 عندما تكون قيمة x تساوي 10

2.3.7: الخطأ المعياري Standard Error:

وهو معيار مهم يقيس كفاءة نموذج الإنحدار المقدر أي مقدار دقة تمثيل النموذج المقدر للعلاقة بين المتغيرين على ضوء البيانات المتاحة لهما وكلما كانت قيمة الخطأ المعياري صغيرة دلَ ذلك على كفاءة النموذج المقدر والعكس صحيح، وعادة تكون الفائدة من هذا المعيار عندما تكون لدينا مقارنة بين نموذجين أو أكثر حول نفس الظاهرتين في دراستين مستقلتين أو أكثر. ويمكن حسابه باستخدام القانون الآتي:

$$S_{y/x} = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-2}} \quad \cdots \quad (7.11)$$

مثال (4.7)**: البيانات التالية تمثل الكمية المعروضة y_i من سلعة معينة وسعر الوحدة الواحدة منها x_i**

الكمية المعروضة y_i	**6**	**8**	**4**	**10**	**13**	**5**	**3**
سعر السلعة x_i	**3**	**4**	**2**	**6**	**7**	**4**	**2**

والمطلوب مايلي:

1- تقدير النموذج الخطي بين الكمية المعروضة (معتمد) وسعرها (مستقل).

2- قدر معامل التحديد وفسر معناه؟

3- قدر كفاءة النموذج من خلال حساب الخطأ المعياري.

الحل:

1- لتقدير النموذج الخطي البسيط يتم تكوين الجدول الآتي:

التسلسل	y_i	x_i	$x_i y_i$	x_i^2	y_i^2	$(y_i - \hat{y}_i)$	$(y_i - \hat{y}_i)^2$
1	6	3	18	9	36	0.7727	0.6
2	8	4	32	16	64	1	1
3	4	2	8	4	16	0.5455	0.3
4	10	6	60	36	100	-0.5455	0.3
5	13	7	91	49	169	0.6818	0.46
6	5	4	20	16	25	-2	4
7	3	2	6	4	9	0.4546-	0.21
المجموع	49	28	235	**134**	**419**	**0**	**6.86**

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{7} y_i}{7} = \frac{49}{7} = 7$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{7} x_i}{7} = \frac{28}{7} = 4$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\cdot\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2} = \frac{235-(7)(4)(7)}{134-(7)(4)^2} = \frac{39}{22} = 1.7727$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 7-(1.7727)(4) = -0.0908$$

وهذا يعني أن نموذج الإنحدار الخطي المقدر هو كمايأتي:

$$\hat{y}_i = -0.0908 + 1.7727\ x_i$$

يمكن تقدير معامل التحديد كمايأتي:

يتم أولاً حساب قيم $\hat{y}_i$ من خلال النموذج المقدر في المطلوب الأول ومن ثم حساب البواقي من خلال طرحها من قيم y_i ومن ثم تربيعها وجمعها كما في الجدول أعلاه. وعلى هذا الأساس نطبق الصيغة (7.9) وكماياتي:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n} y_i^2 - n\,\bar{y}^2}$$

$$= 1 - \frac{6.86}{419-(7)(7)^2} = 1 - \frac{6.86}{76} = 1 - 0.0903 = 0.9097$$

وهذا يعني أن 90.97% من التغيرات الحاصلة في الكمية المعروضة سببها سعر الوحدة الواحدة منها والباقي 9.03% سببها متغيرات أخرى غير مدروسة.

تقدير كفاءة النموذج من خلال حساب الخطأ المعياري هو كمايأتي:

$$S_{y/x} = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-2}} = \sqrt{\frac{6.86}{7-2}} = \sqrt{1.372} = 1.1713$$

وهذا يعني أن الخطأ المعياري الذي يقيس كفاءة النموذج هو 1.1713

مثال (5.7)**: إذا توفرت لديك المعلومات الآتية:**

$n = 15$	$S_y^2 = 50$	$R^2 = 0.75$

جد الخطأ المعياري؟

الحل: من تباين المتغير المعتمد يمكن إيجاد مجموع مربعات إنحرافات المتغير المعتمد عن الوسط الحسابي وكمايأتي:

$$S_y^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1} \Rightarrow 50 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{15-1} \Rightarrow \sum_{i=1}^{n}(y_i - \bar{y})^2 = (50)\cdot(14) = 700$$

كذلك لدينا:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \Rightarrow 0.75 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{700}$$

$$\Rightarrow 1 - 0.75 = \frac{\sum_{i=1}^{n} e_i^2}{700} \Rightarrow (0.25)\cdot(700) = \sum_{i=1}^{n} e_i^2$$

$$\therefore \quad \sum_{i=1}^{n} e_i^2 = 175$$

فإن الخطأ المعياري هو كمايأتي:

$$S_{y/x} = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-2}} = \sqrt{\frac{175}{15-2}} = \sqrt{13.4615} = 3.669$$

4.7: الإنحدار الخطي المتعدد Multiple Linear Regression:

يستخدم الإنحدار الخطي المتعدد في عملية تقدير العلاقة الخطية بين المتغير المعتمد وعدة متغيرات مستقلة (أكثر من متغير مستقل واحد) ربما تؤثر مجتمعةً في المتغير المعتمد. وهنا سوف يتم التركيز على متغيرين مستقلين فقط، أما أكثر من ذلك فهو خارج نطاق هدف هذا الكتاب. لذلك فإن النموذج المتعدد ذات متغيرين مستقلين هو كمايأتي:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i \qquad \cdots \quad (7.12)$$

يمثل y_i مشاهدات المتغير المعتمد (مثل الإنفاق الشهري للعائلة) في حين يمثل x_{i1} مشاهدات المتغير المستقل الأول (مثل الدخل الشهري للأسرة) ويمثل x_{i2} مشاهدات المتغير المستقل الثاني (مثل عدد أفراد أسرة)، في حين يمثل ε_i الخطأ العشوائي (يجب وجوده في النموذج لأنه ليس لدينا علاقة تامة أو محددة Deterministic بين المتغير المعتمد والمتغيرين المستقلين) والذي يمثل التغيرات الحاصلة في المتغير المعتمد نتيجة تأثيرات عدة متغيرات أخرى (عدى المتغيرين المستقلين) غير مدروسة. وأن β_0 ، β_1 و β_2 تمثل معلمات نموذج الإنحدار الخطي المتعدد للمجتمع وتستخدم إحدى طرائق التقدير في تقدير معلمات خط الانحدار المتعدد للعينة مثل طريقة المربعات الصغرى الإعتيادية (OLS)، أي تقدير معلمات النموذج الآتي:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} \qquad \cdots \quad (7.13)$$

$\hat{y}_i$ تمثل القيم المقدرة للمتغير المعتمد من خلال النموذج المقدر أعـلاه وأن $\hat{\beta}_0$ ، $\hat{\beta}_1$ و $\hat{\beta}_2$ تمثل معلمات النموذج المقدرة من العينة والتي يمكن حسابها من خلال الصيغ الآتية:

$$\hat{\beta}_1 = \frac{\left(\sum_{i=1}^{n} X_{i1} Y_i\right)\left(\sum_{i=1}^{n} X_{i2}^2\right) - \left(\sum_{i=1}^{n} X_{i2} Y_i\right)\left(\sum_{i=1}^{n} X_{i1} X_{i2}\right)}{\left(\sum_{i=1}^{n} X_{i1}^2\right)\left(\sum_{i=1}^{n} X_{i2}^2\right) - \left(\sum_{i=1}^{n} X_{i1} X_{i2}\right)^2} \quad \cdots \quad (7.14)$$

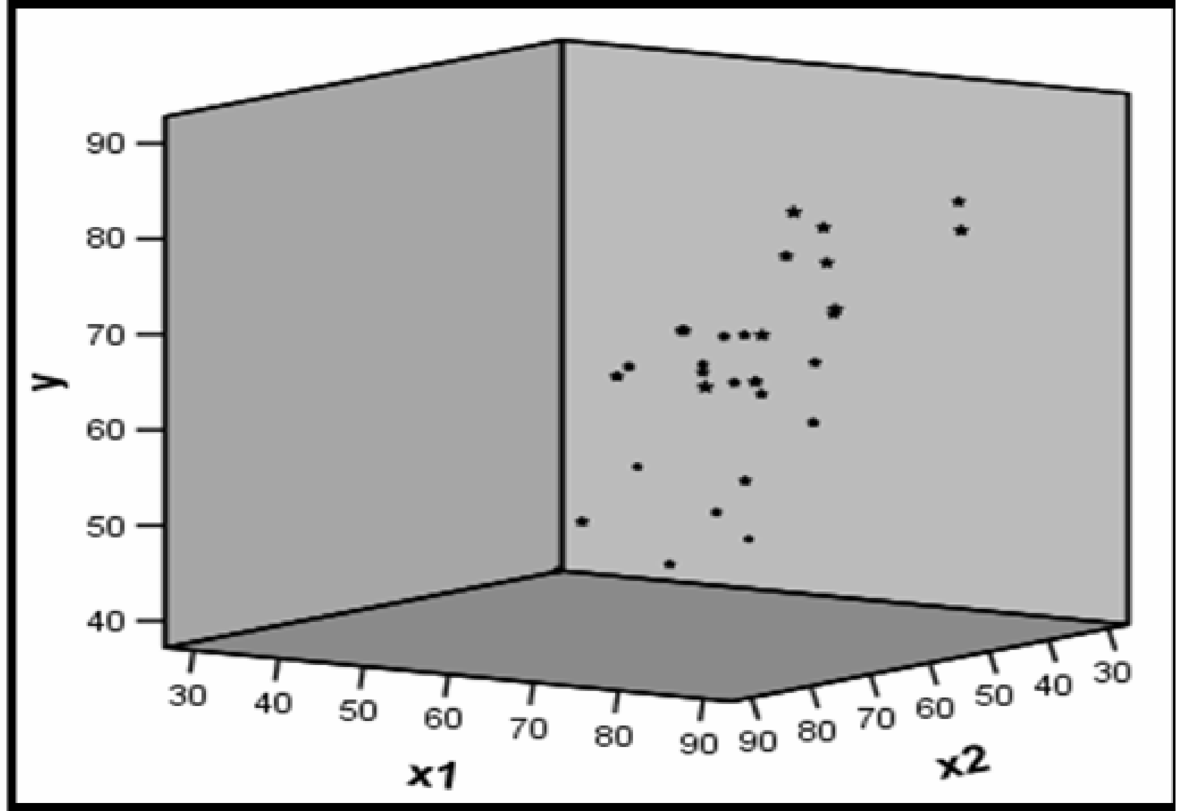
$$\hat{\beta}_2 = \frac{\left(\sum_{i=1}^{n} X_{i2} Y_i\right)\left(\sum_{i=1}^{n} X_{i1}^2\right) - \left(\sum_{i=1}^{n} X_{i1} Y_i\right)\left(\sum_{i=1}^{n} X_{i1} X_{i2}\right)}{\left(\sum_{i=1}^{n} X_{i1}^2\right)\left(\sum_{i=1}^{n} X_{i2}^2\right) - \left(\sum_{i=1}^{n} X_{i1} X_{i2}\right)^2} \quad \cdots \quad (7.15)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2 \quad \cdots \quad (7.16)$$

حيث أن قيم X_{i2}, X_{i1}, Y_i تمثل إنحرافات القيم عن الوسط الحسابي، أي أن:

$$Y_i = y_i - \bar{y}, \quad X_{i1} = x_{i1} - \bar{x}_1, \quad X_{i2} = x_{i2} - \bar{x}_2$$

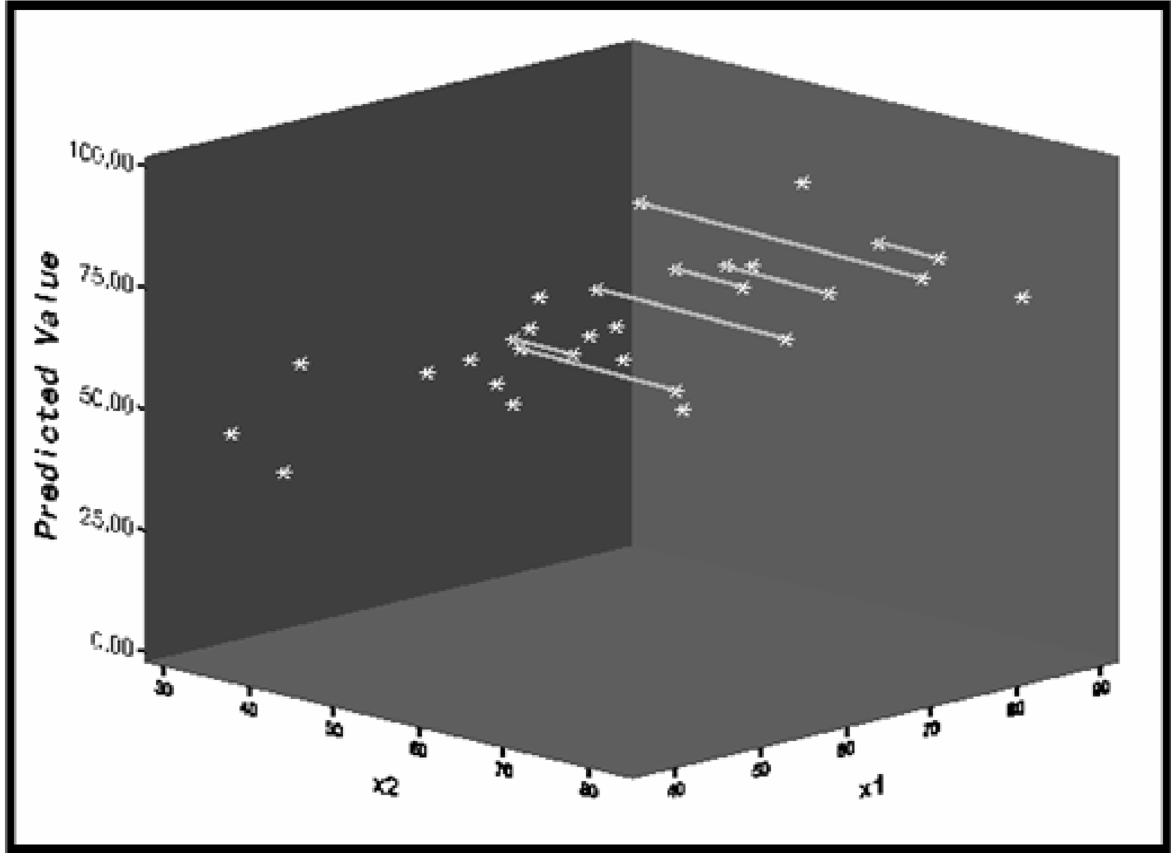
أن الشكل الإنتشاري يكون ثلاثي الأبعاد كما نلاحظ من خلال الشكل التالي لبيانات تمثل 30 مشاهدة لمتغير معتمد مع متغيرين مستقلين:



الشكل (7.5)

يبين الشكل الإنتشاري بين المتغير المعتمد مقابل متغيرين مستقلين

إن نموذج الإنحدار المتعدد ذات المتغيرين المستقلين يمثل أفضل مستوى إنحدار في الفضاء الثلاثي الأبعاد $\left(y, x_1, x_2\right)$ الذي يعبر عن العلاقة مابين المتغير المعتمد من جهة مع المتغيرين المستقلين من جهة أخرى لذلك فإن الشكل الإنتشاري (الشكل (7.5)) يمكن تقدير النموذج المتعدد الملائم له ومن ثم رسم هذا النموذج كما يوضحه الشكل الآتي:



الشكل (7.6)

يبين مستوى الإنحدار بين المتغير المعتمد مقابل متغيرين مستقلين

إن $\hat{\beta}_0$ تمثل نقطة تقاطع مستوى الإنحدار مع المحور y وأن $\hat{\beta}_1$ تمثل معامل إنحدار y على x_1 بثبوت أثر x_2 وأن $\hat{\beta}_2$ تمثل معامل إنحدار y على x_2 بثبوت أثر x_1 وهذا يعني أن $\hat{\beta}_1$ تمثل مقدار تغيرy عند تغير x_1 بوحدة واحدة عند ثبات x_2 وأن $\hat{\beta}_2$ تمثل مقدار تغير y عند تغير x_2 بوحدة واحدة عند ثبات x_1 . وإذا كانت قيمتا معاملي الإنحدار موجبة دلَ ذلك على وجود علاقة طردية بين المتغير المعتمد والمتغيرين المستقلين أما إذا كانت سالبتان دلَ ذلك على وجود علاقة عكسية بين المتغير المعتمد والمتغيرين المستقلين في حين إذا كانت الأولى موجبة والثانية سالبة دلَ ذلك على

وجود علاقة طردية بين المتغير المعتمد والمتغير المستقل الأول وعلى وجود علاقة عكسية بين المتغير المعتمد والمتغير المستقل الثاني (وعكس هذه الحالة صحيح) وأخيراً إذا كانت قيمتا معاملي الإنحدار مساوية للصفر فذلك يعني أنه لاتوجد علاقة بين المتغير المعتمد والمتغيرين المستقلين وأن التغيرات الحاصلة في المتغير المعتمد لاتفسرها تلك المتغيرات المستقلة بل هنالك متغيرات أخرى غير مدروسة هي المسببة لتلك التغيرات في المتغير المعتمد.

مثال (6.7)**: البيانات التالية تمثل نتائج إستطلاع رأي 10 أسر فيما يخص متوسط إنفاقها الشهري على الملابس y_i ، إنفاقها الكلي x_{i1} ، متوسط أسعار مجموعة من الملابس x_{i2} (بآلآف الدنانير):**

الإنفاق على الملابس y_i	الإنفاق الكلي x_{i1}	متوسط أسعار الملابس x_{i2}
20	105	15
35	110	12
18	90	25
24	108	16
36	150	10
50	200	4
45	160	4
12	90	30
15	100	22
20	100	14

والمطلوب مايأتي:

1- **تقدير دالة الإنحدار المتعدد للأنفاق على الملابس، وفسر من خلاله علاقة الإنفاق على الملابس مع الإنفاق الكلي ومتوسط أسعار الملابس.**
2- **قدر قيمة الإنفاق على الملابس عندما يكون الإنفاق الكلي 250 أف دينار**

ومتوسط أسعار الملابس 2 ألف دينار.

الحل:

1- لتقدير دالة الإنحدار المتعدد للإنفاق على الملابس يجب أولاً حساب الأوساط الحسابية للمتغيرات الثلاث وكما يأتي:

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{10} y_i}{10} = \frac{275}{10} = 27.5$$

$$\overline{x}_1 = \frac{\sum_{i=1}^{n} x_{i1}}{n} = \frac{\sum_{i=1}^{10} x_{i1}}{10} = \frac{1213}{10} = 121.3$$

$$\overline{x}_2 = \frac{\sum_{i=1}^{n} x_{i2}}{n} = \frac{\sum_{i=1}^{10} x_{i2}}{10} = \frac{152}{10} = 15.2$$

والتي يتم من خلالها حساب إنحرافات القيم عن وسطها الحسابي، أي أن:

$$Y_i = y_i - \overline{y}, \quad X_{i1} = x_{i1} - \overline{x}_1, \quad X_{i2} = x_{i2} - \overline{x}_2$$

تم تلخيص هذه الإنحرافات مع بعض العمليات الجبرية المطلوبة من خلال الجدول الآتي:

$X_{i2}Y$	$X_{i1}Y_i$	X_{i2}^2	X_{i1}^2	$X_{i1}X_{i2}$	X_{i2}	X_{i1}	Y_i
1.5	122.25	0.04	265.69	3.26	0.2-	16.3-	7.5-
24-	84.75-	10.24	127.69	36.16	3.2-	11.3-	7.5
93.1-	297.35	96.04	979.69	306.74-	9.8	31.3-	9.5-
2.8-	46.55	0.64	176.89	10.64-	0.8	13.3-	3.5-
44.2-	243.95	27.04	823.69	149.24-	5.2-	28.7	8.5
252-	1770.75	125.44	6193.69	881.44-	11.2-	78.7	22.5
196-	677.25	125.44	1497.69	433.44-	11.2-	38.7	17.5

229.4-	485.15	219.04	979.69	463.24-	14.8	31.3-	15.5-
85-	266.25	46.24	453.69	144.84-	6.8	21.3-	12.5-
9	159.75	1.44	453.69	25.56	1.2-	21.3-	7.5-
916-	3984.5	651.6	11952.1	2324.6-	0	0	0

وعلى هذا الأساس يمكن تقدير معلمات الإنحدار المتعدد وكمايأتي:

$$\hat{\beta}_1 = \frac{\left(\sum_{i=1}^{n} X_{i1} Y_i\right)\left(\sum_{i=1}^{n} X_{i2}^2\right) - \left(\sum_{i=1}^{n} X_{i2} Y_i\right)\left(\sum_{i=1}^{n} X_{i1} X_{i2}\right)}{\left(\sum_{i=1}^{n} X_{i1}^2\right)\left(\sum_{i=1}^{n} X_{i2}^2\right) - \left(\sum_{i=1}^{n} X_{i1} X_{i2}\right)^2}$$

$$= \frac{(3984.5)(651.6) - (-916)(-2324.6)}{(11952.1)(651.6) - (-2324.6)^2} = \frac{466966.6}{2384223.2} = 0.1959$$

$$\hat{\beta}_2 = \frac{\left(\sum_{i=1}^{n} X_{i2} Y_i\right)\left(\sum_{i=1}^{n} X_{i1}^2\right) - \left(\sum_{i=1}^{n} X_{i1} Y_i\right)\left(\sum_{i=1}^{n} X_{i1} X_{i2}\right)}{\left(\sum_{i=1}^{n} X_{i1}^2\right)\left(\sum_{i=1}^{n} X_{i2}^2\right) - \left(\sum_{i=1}^{n} X_{i1} X_{i2}\right)^2}$$

$$= \frac{(-916)(11952.1) - (3984.5)(-2324.6)}{(11952.1)(651.6) - (-2324.6)^2} = \frac{-1685754.9}{2384223.2} = -0.707$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2$$

$$= 27.5 - (0.1959)(121.3) - (-0.707)(15.2) = 14.4837$$

لذلك فإن دالة الإنحدار المتعدد للإنفاق على الملابس هي كمايأتي:

$$\hat{y}_i = 14.4837 + 0.1959\, x_{i1} - 0.707\, x_{i2}$$

مـن خـلال دالـة الإنفـاق عـلى الملابـس المقـدرة نلاحـظ أن هنالـك علاقـة طرديـة بـين الإنفـاق عـلى الملابـس والإنفـاق الكـلي لإن قيمـة $\hat{\beta}_1$ موجبـة في حـين هنالـك علاقـة عكـسية

بين الإنفاق على الملابس ومتوسط أسعارها لإن قيمة $\hat{\beta}_2$ سالبة.

-2 **لتقدير قيمة الإنفاق على الملابس $\hat{y}_0$ عندما $x_1 = 250$ و $x_2 = 2$ لدينا مايأتي:**

$$\hat{y}_i = 14.4837 + 0.1959\, x_{i1} - 0.707\, x_{i2}$$
$$\hat{y}_0 = 14.4837 + 0.1959\,(250) - 0.707\,(2) = 62.045$$

وهذا يعني أن مقدار الإنفاق على الملابس هو 62.045 ألف دينار عندما يكون الإنفاق الكلي هو 250 ألف دينار ومتوسط أسعار الملابس هو 2 ألف دينار.

ملاحظة: إن مجموع البواقي للإنحدار الخطي المتعدد أيضاً تساوي صفر.

1.4.7: معامل التحديد للإنحدار الخطي المتعدد:

معامل التحديد للإنحدار الخطي المتعدد هو المقدار الذي يفسر التغيرات الحاصلة في المتغير المعتمد y نتيجة تأثره بعدة متغيرات مستقلة مجتمعةً (أكثر من متغير مستقل واحد) أو هو درجة مساهمة المتغيرات المستقلة في التغيرات الحاصلة في y والذي يمكن حسابه من خلال إيجاد مربع معامل الإرتباط المتعدد بين y والمتغيرات المستقلة مجتمعةً وإذا كان هنالك متغيرين مستقلين فقط فإن معامل التحديد هو كمايأتي:

$$R^2_{y \cdot x_1 x_2} = \left(r_{y \cdot x_1 x_2}\right)^2 \qquad \cdots \quad (7.17)$$

أو يمكن حسابه من خلال الصيغة الآتية:

$$R^2_{y \cdot x_1 x_2} = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} Y_i^2} \qquad \cdots \quad (7.18)$$

أو يمكن حسابه من خلال الصيغة الآتية:

$$R^2_{y\cdot x_1 x_2} = \frac{\sum_{i=1}^{n} \left(\hat{y}_i - \overline{y}\right)^2}{\sum_{i=1}^{n} Y_i^2} \qquad \cdots \quad (7.19)$$

أو يمكن حسابه من خلال الصيغة الآتية:

$$R^2_{y\cdot x_1 x_2} = \frac{\hat{\beta}_1 \cdot \sum_{i=1}^{n} X_{i1} Y_i + \hat{\beta}_2 \cdot \sum_{i=1}^{n} X_{i2} Y_i}{\sum_{i=1}^{n} Y_i^2} \qquad \cdots \quad (7.20)$$

مثال (7.7)**: جد معامل التحديد للمثال** (6.7) **ومن خلاله جد معامل الإرتباط المتعدد بين الإنفاق على الملابس مع الإنفاق الكلي ومتوسط أسعار الملابس.**

الحل: نحتاج أولاً حساب مجموع إنحرافات القيم عن وسطها الحسابي للمتغير y وكمايأتي:

$$\sum_{i=1}^{n} Y_i^2 = (-7.5)^2 + (7.5)^2 + \cdots + (-7.5)^2 = 1552.5$$

ومن خلال معلومات المثال السابق يمكن تطبيق الصيغة (7.20) وكمايأتي:

$$R^2_{y\cdot x_1 x_2} = \frac{\hat{\beta}_1 \cdot \sum_{i=1}^{n} X_{i1} Y_i + \hat{\beta}_2 \cdot \sum_{i=1}^{n} X_{i2} Y_i}{\sum_{i=1}^{n} Y_i^2}$$

$$= \frac{(0.1959)(3984.5) + (-0.707)(-916)}{1552.5} = \frac{1428.1756}{1552.5} = 0.92$$

وهذا يعني أن 92% من التغيرات الحاصلة في الإنفاق على الملابس يفسره الإنفاق الكلي ومتوسط أسعار الملابس.

وبما أن العلاقة طردية بين المتغير المعتمد والمتغيرات المستقلة بشكل عام (لأنه يجب الإنتباه على إتجاه العلاقة لأنها تكون غير واضحة باستخدام هذه الطريقة) يمكن إيجاد معامل الإرتباط المتعدد من خلال أخذ الجذر التربيعي لمعامل التحديد وكمايأتي:

$$r_{y\bullet x_1 x_2} = \sqrt{R^2_{y\bullet x_1 x_2}} = \sqrt{0.92} = 0.9592$$

وهذا يعني هنالك علاقة طردية قوية تساوي 95.92% بين الإنفاق على الملابس مع الإنفاق الكلي ومتوسط أسعار الملابس.

2.4.7: الخطأ المعياري للإنحدار الخطي المتعدد:

وهو يقيس كفاءة نموذج الإنحدار الخطي المتعدد أي مقدار دقة تمثيل النموذج المقدر للعلاقة بين المتغير المعتمد مع المتغيرات المستقلة (هنا أيضاً سنكتفي بمتغيرين مستقلين فقط) على ضوء البيانات المتاحة وكلما كانت قيمة الخطأ المعياري صغيرة دلَ ذلك على كفاءة النموذج المقدر والعكس صحيح، وهو لايختلف كثيراً عن الخطأ المعياري المقدر للنموذج الخطي البسيط. ويمكن حسابه باستخدام القانون الآتي:

$$S_{y/x_1 x_2} = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-3}} \qquad \cdots \quad (7.21)$$

مثال (8.7)**: جد الخطأ المعياري من المعلومات المتوفرة من المثال** (6.7) **والمثال** (7.7).

الحل: لدينا مايلي من المعلومات:

$$n = 10 \quad \sum_{i=1}^{n} Y_i^2 = 1552.5 \quad R^2_{y\bullet x_1 x_2} = 0.92$$ لذلك فإن:

$$R^2_{y.x_1x_2} = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} Y_i^2} \Rightarrow 0.92 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{1552.5} \Rightarrow 0.08 = \frac{\sum_{i=1}^{n} e_i^2}{1552.5}$$

$$\Rightarrow \sum_{i=1}^{n} e_i^2 = (0.08)(1552.5) = 124.2$$

وعلى هذا الأساس يمكن الحصول على الخطأ المعياري وكمايأتي:

$$S_{y/x_1x_2} = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-3}} = \sqrt{\frac{124.2}{10-3}} = \sqrt{17.7429} = 4.2122$$

ملاحظة: من الناحية العملية دراسة العلاقة بين متغيرين أحدهما معتمد والآخر مستقل هي غير واقعية لأنه لايوجد متغير يتأثر بمتغير واحد فقط وإنما بعدة متغيرات أخرى وأنه كلما إزداد عدد المتغيرات المستقلة فذلك يعني إضعاف الخطأ العشوائي بسبب أخذ جزء أكبر من مكوناته وتشخيصها في النموذج وهذه العملية تؤثر في تقديرات معلمات النموذج المقدر وفي المعايير المحسوبه له مثل معامل التحديد والخطأ المعياري.

مثال : (9.7)**أفرض لديك البيانات الآتية:**

	y	x1	x2
1	2	3	1
2	5	4	0
3	4	5	1
4	0	3	0
5	6	7	0
6	1	2	4

والمطلوب تكوين جدول مقارنة بين نموذج الإنحدار الخطي البسيط للمتغيرات y و x_1 ونموذج الإنحدار الخطي المتعدد للمتغيراتy و x_1 و x_2 من حيث المعلمات المقدرة ومعاملي الإرتباط والتحديد والخطأ المعياري.

الحل: يتم أولاً حساب الأوساط الحسابية للمتغيرات الثلاث وكمايأتي:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{6} y_i}{6} = \frac{18}{6} = 3$$

$$\bar{x}_1 = \frac{\sum_{i=1}^{n} x_{i1}}{n} = \frac{\sum_{i=1}^{6} x_{i1}}{6} = \frac{24}{6} = 4$$

$$\bar{x}_2 = \frac{\sum_{i=1}^{n} x_{i2}}{n} = \frac{\sum_{i=1}^{6} x_{i2}}{6} = \frac{6}{6} = 1$$

والتي يتم من خلالها حساب إنحرافات القيم عن وسطها الحسابي، أي أن:

$$Y_i = y_i - \bar{y}, \quad X_{i1} = x_{i1} - \bar{x}_1, \quad X_{i2} = x_{i2} - \bar{x}_2$$

تم تلخيص هذه الإنحرافات مع بعض العمليات الجبرية المطلوبة من خلال الجدول الآتي:

$X_{i2}Y$	$X_{i1}Y_i$	$X_{i1}X_{i2}$	X_{i2}^2	X_{i1}^2	Y_i^2	X_{i2}	X_{i1}	Y_i	التسلسل
0	1	0	0	1	1	0	1-	1-	1
2-	0	0	1	0	4	1-	0	2	2
0	1	0	0	1	1	0	1	1	3
3	3	1	1	1	9	1-	1-	3-	4
3-	9	3-	1	9	9	1-	3	3	5
6-	4	6-	9	4	4	3	2-	2-	6
8-	18	8-	12	16	28	0	0	0	المجموع

وعلى هذا الأساس يمكن تقدير معلمات الإنحدار الخطي البسيط وكمايأتي:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)(y_i - \bar{y})}{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2} = \frac{\sum_{i=1}^{n} X_{i1} Y_i}{\sum_{i=1}^{n} X_{i1}^2} = \frac{18}{16} = 1.125$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 = 3 - (1.125)(4) = -1.5$$

وهذا يعني أن النموذج المقدر هو كالآتي:

$$\hat{y}_i = -1.5 + 1.125\, x_{i1}$$

وهذا يعني أن هنالك علاقة طردية بين y و x_1 لإن معامل الإنحدار له قيمة موجبـة. وأن زيـادة x_1 بمقدار وحدة واحدة ستؤدي إلى زيادة في y مقدارها 1.125

- معامل الإرتباط الخطي البسيط بينy و x_1 هو كمايأتي:

$$r_{x_1 y} = \frac{\sum_{i=1}^{n} (x_{i1} - \bar{x}_1)(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_{i1} - \bar{x}_1)^2 . \sum_{i=1}^{n} (y_i - \bar{y})^2}} = \frac{\sum_{i=1}^{n} X_{i1} Y_i}{\sqrt{\sum_{i=1}^{n} X_{i1}^2 \cdot \sum_{i=1}^{n} Y_i^2}}$$

$$= \frac{18}{\sqrt{(16)(28)}} = \frac{18}{21.166} = 0.85$$

وهذا يعني أن هنالك علاقة طردية قوية تساوي 85% بين المتغير المعتمد والمستقل x_1

- معامل التحديد للنموذج الخطي البسيط المقدر آنفاً هو كمايأتي:

$$R^2 = (r_{x_1 y})^2 = (0.85)^2 = 0.7225$$

وهذا يعني أن72.25% من التغـيرات الحاصـلة في y يفـسرها المتغـير المـستقل x_1 والبـاقي 27.75% تفسرها متغيرات مستقلة أخرى غير مدروسة.

- تقدير الخطأ المعياري للنموذج الخطي البسيط المقدر آنفاً هو كمايأتي:

باستخدام النموذج الخطي البسيط يمكن حساب القيم المقدرة للمتغير المعتمد $\hat{y}_i$ وكمايأتي:

$$\hat{y}_i = -1.5 + 1.125\, x_{i1}$$
$$\hat{y}_1 = -1.5 + 1.125\, x_{11} = -1.5 + 1.125\,(3) = 1.875$$
$$\hat{y}_2 = -1.5 + 1.125\, x_{21} = -1.5 + 1.125\,(4) = 3$$
$$\vdots$$
$$\hat{y}_6 = -1.5 + 1.125\, x_{61} = -1.5 + 1.125\,(2) = 0.75$$

ولحساب قيم البواقي نستخدم الصيغة (7.3):

$$e_i = y_i - \hat{y}_i$$
$$e_1 = y_1 - \hat{y}_1 = 2 - 1.875 = 0.125$$
$$e_2 = y_2 - \hat{y}_2 = 5 - 3 = 2$$
$$\vdots$$
$$e_6 = y_6 - \hat{y}_6 = 1 - 0.75 = 0.25$$

وعلى هذا الأساس فإن مجموع مربعات البواقي هو كمايأتي:

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{6} e_i^2 = e_1^2 + e_2^2 + \cdots + e_6^2$$
$$= (0.125)^2 + (2)^2 + \cdots + (0.125)^2 = 7.75$$

وأن الخطأ المعياري هو كمايأتي:

$$S_{y/x_1} = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-2}} = \sqrt{\frac{7.75}{6-2}} = 1.3919$$

يمكن تقدير معلمات الإنحدار الخطي المتعدد كمايأتي:

$$\hat{\beta}_1 = \frac{\left(\sum_{i=1}^{n} X_{i1} Y_i\right)\left(\sum_{i=1}^{n} X_{i2}^2\right) - \left(\sum_{i=1}^{n} X_{i2} Y_i\right)\left(\sum_{i=1}^{n} X_{i1} X_{i2}\right)}{\left(\sum_{i=1}^{n} X_{i1}^2\right)\left(\sum_{i=1}^{n} X_{i2}^2\right) - \left(\sum_{i=1}^{n} X_{i1} X_{i2}\right)^2}$$

$$= \frac{(18)(12) - (-8)(-8)}{(16)(12) - (-8)^2} = \frac{152}{128} = 1.1875$$

$$\hat{\beta}_2 = \frac{\left(\sum_{i=1}^{n} X_{i2} Y_i\right)\left(\sum_{i=1}^{n} X_{i1}^2\right) - \left(\sum_{i=1}^{n} X_{i1} Y_i\right)\left(\sum_{i=1}^{n} X_{i1} X_{i2}\right)}{\left(\sum_{i=1}^{n} X_{i1}^2\right)\left(\sum_{i=1}^{n} X_{i2}^2\right) - \left(\sum_{i=1}^{n} X_{i1} X_{i2}\right)^2}$$

$$= \frac{(-8)(16) - (18)(-8)}{(16)(12) - (-8)^2} = \frac{16}{128} = 0.125$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2$$

$$= 3 - (1.1875)(4) - (0.125)(1) = -1.875$$

لذلك فإن نموذج الإنحدار المتعدد هو كمايأتي:

$$\hat{y}_i = -1.875 + 1.1875\, x_{i1} + 0.125\, x_{i2}$$

وهذا يعني أن هنالك علاقة طردية بين المتغير المعتمـد والمتغـيرين المـستقلين لأن قيمتـا معـاملي الإنحدار موجبتان.

- معامل التحديد للنموذج الخطي المتعدد بين y مع x_1 و x_2 هو كمايأتي:

$$R^2_{y.x_1x_2} = \frac{\hat{\beta}_1 \cdot \sum_{i=1}^{n} X_{i1} Y_i + \hat{\beta}_2 \cdot \sum_{i=1}^{n} X_{i2} Y_i}{\sum_{i=1}^{n} Y_i^2}$$

$$= \frac{(1.1875)(18) + (0.125)(-8)}{28} = \frac{20.375}{28} = 0.7277$$

وهذا يعني أن%72.77 من التغيرات الحاصلة في y يفسرها المتغيرين المستقلين x_1 و x_2 والباقي %27.23 تفسرها متغيرات مستقلة أخرى غير مدروسة.

- **معامل الإرتباط الخطي المتعدد بين y مع x_1 و x_2 هو كمايأتي:**

$$r_{y \cdot x_1 x_2} = \sqrt{R^2_{y \cdot x_1 x_2}} = \sqrt{0.7277} = 0.8531$$

وهذا يعني أن هنالك علاقة طردية قوية تساوي %85.31 بين المتغير المعتمد مع المتغيرين المستقلين x_1 و x_1.

وأن الخطأ المعياري هو كمايأتي:

$$R^2_{y \cdot x_1 x_2} = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} Y_i^2} \Rightarrow 0.7277 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{28} \Rightarrow 0.2723 = \frac{\sum_{i=1}^{n} e_i^2}{28}$$

$$\Rightarrow \sum_{i=1}^{n} e_i^2 = (0.2723)(28) = 7.6244$$

وعلى هذا الأساس يمكن الحصول على الخطأ المعياري وكمايأتي:

$$S_{y/x_1 x_2} = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-3}} = \sqrt{\frac{7.6244}{6-3}} = \sqrt{2.5415} = 1.5942$$

والآن يمكن تكوين جدول المقارنة بين النموذجين وكمايأتي:

المعلمة أو المعيار	نموذج الإنحدار الخطي البسيط	نموذج الإنحدار الخطي المتعدد
$\hat{\beta}_0$	1.5-	1.875-
$\hat{\beta}_1$	1.125	1.1875
$\hat{\beta}_2$	---	0.125
معامل الإرتباط	0.85	0.8531
معامل التحديد	0.7225	0.7277
الخطأ المعياري	1.3919	1.5942

نلاحظ من الجدول أعلاه أن قيم المعلمات أختلفت نتيجة إضافة المتغير المستقل الثاني إلى النموذج الخطي البسيط مع بقاء إشارة معامل الإنحدار كماهي. وأن هنالك زيادة محدودة في معاملي الإرتباط والتحديد في النموذج الخطي المتعدد على النموذج الخطي البسيط. في حين أن الخطأ المعياري قد أزداد دلالةً على عدم كفاءة النموذج الخطي المتعدد مقارنةً بنموذج الخطي البسيط. وذلك يعني إمكانية الإعتماد على النموذج الخطي البسيط لأن قيمة الخطأ المعياري كانت أقل (أفضل)، وذلك لعدم أهمية المتغير الثاني المضاف إلى النموذج.

ملاحظة: عادة تنخفض قيمة الخطأ المعياري في النموذج الخطي المتعدد إذا كان المتغير المضاف إلى النموذج الخطي البسيط ذات أهمية كبيرة في تفسير التغيرات الحاصلة في المتغير المعتمد ويمكن معرفة ذلك عندما نلاحظ أن هنالك زيادة ملحوظة في قيمتا معاملي الإرتباط والتحديد (هنالك زيادة معنوية ويمكن التأكد منها من خلال الإختبارات الإحصائية).

تمارين الفصل السابع

7-1: ما المقصود بمعامل الإنحدار الخطي البسيط ؟

7-2: إشرح العلاقة مابين الإنحدار الخطي البسيط ومعامل الإنحدار.

7-3: في تجربة زراعية لدراسة أثر زيادة كمية السماد العضوي على كمية محصول الحنطة تم الحصول على النتائج الآتية:

كمية السماد	14	10	3	9	11	15	8	13	17	19	7	8	9	12
كمية المحصول	9	7	2	5	8	11	6	8	13	15	4	5	6	7

والمطلوب مايأتي:

1- **أرسم إنتشار البيانات التي تمثل كمية السماد مقابل كمية المحصول.**
2- **قدر النموذج الخطي بين كمية السماد وكمية المحصول.**
3- **ماهي الكمية المتوقعة من المحصول إذا كانت كمية السماد المعطاة تساوي 20**
4- **جد معامل التحديد وفسر معناه.**
5- **جد معامل الإرتباط الخطي البسيط بين كمية السماد وكمية محصول الحنطة.**

7-4: سجلت إحدى دوائر الأنواء الجوية في إقليم كوردستان البيانات التالية لحالة الطقس ولثمانية أيام متتالية ممطرة عن كل من درجة الحرارة وكمية المطر الساقطة:

درجة الحرارة	2-	1	1-	3	2	3	1	2
كمية المطر	5	4	6	7	5	4	3	2

والمطلوب مايأتي:

1- تحديد المتغير المستقل x والمتغير المعتمد y.

2- تقدير نموذج الإنحدار الخطي البسيط y/x .

3- رسم الشكل الإنتشاري لدرجات الحرارة مقابل كمية المطر مع خط الإنحدار المقدر.

4- حساب معامل الإرتباط الخطي البسيط وتفسير معناه.

5- حساب معامل التحديد وتفسير معناه.

6- حساب الخطأ المعياري للنموذج المقدر في المطلوب الثاني.

7- ما قيمة y إذا كانت x = 8 ،9، 10

5-7: البيانات التالية تمثل الكمية المطلوبة من سلعة معينة y، الدخل الشهري x_1 وسعر الوحدة الواحدة من هذه السلعة x_2 لعينة تتألف من 12 عائلة.

y	2	5	6	4	5	4	9	8	10	6	13	7
x_1	90	110	120	95	105	110	135	125	145	115	150	118
x_2	10	7	6	4	5	2	5	5	4	3	4	3

والمطلوب مايأتي:

1- تقدير نموذج الإنحدار المتعدد للكمية المطلوبة من هذه السلعة بالإعتماد على الدخل الشهري وسعر الوحدة الواحدة منها.

2- قدر معامل الإرتباط المتعدد بين y والمتغيرين المستقلين x_1 و x_2 وفسر معناه.

3- أحسب معامل التحديد للنموذج الخطي المتعدد المقدر في المطلوب الأول وفسر معناه.

4- أحسب الخطأ المعياري للنموذج المقدر في المطلوب الأول.

5- قدر الكمية المطلوبة من هذه السلعة عندما $x_1 = 200$ و $x_2 = 12$.

6- قدر نموذج الإنحدار الخطي البسيط للكمية المطلوبة من هذه السلعة بالإعتماد على الدخل الشهري فقط.

7- أحسب معامل الإرتباط الخطي البسيط بين الكمية المطلوبة من هذه السلعة والدخل الشهري.

8- أحسب معامل التحديد للنموذج الخطي البسيط المقدر في المطلوب السادس

9- أحسب الخطأ المعياري للنموذج المقدر في المطلوب السادس.

10- قارن بين النموذجين المقدرين في المطلوب الأول والسادس من خلال الخطأ المعياري.

6-7: إذا توفرت لديك المعلومات الآتية:

$\bar{Y} = 7.14$	$\bar{X}_1 = 9$	$\bar{X}_2 = -1.57$	n=7
$S_y = 6.012$	$S_1 = 6.377$	$S_2 = 4.541$	
$r_{y1} = 0.991$	$r_{y2} = -0.912$	$r_{12} = -0.955$	

والمطلوب مايأتي:

1- تقدير نموذج الإنحدار الخطي البسيط y/x_1.

2- تقدير نموذج الإنحدار الخطي البسيط y/x_2.

3- تقدير نموذج الإنحدار الخطي البسيط x_1/x_2.

4- تقدير نموذج الإنحدار الخطي البسيط x_2/x_1.

5- تقدير نموذج الإنحدار الخطي المتعدد $y/x_1\,x_2$.

6- أرسم الإنتشار مع خط الإنحدار للنماذج المقدرة أعلاه.

7- حساب معامل التحديد للنماذج المقدرة أعلاه.

8- **حساب الخطأ العياري للنماذج المقدرة أعلاه.**

9- **قارن بين كفاءة النماذج المقدرة للمتغيرy من خلال معامل التحديد والخطأ المعياري.**

10- **أثبت حسابياً أن مجموع البواقي للنماذج المقدرة أعلاه يساوي صفر.**

11- **تقدير قيمة** y **عندما** $x_1 = 20$.

12- **تقدير قيمة** y **عندما** $x_2 = 10$.

13- **تقدير قيمة** y **عندما** $x_1 = 20$ **و** $x_2 = 10$.

الملاحق

الملحق A: المصطلحات العلمية

الملحق B: الصيغ الإحصائية

الملحق A

المصطلحات العلمية

عربي	إنكليزي
إحتمالات	Probability
إحصاء	Statistics
إحصاء إستدلالي	Inference Statistical
إحصاء تطبيقي	Applied Statistics
إحصاء وصفي	Descriptive Statistics
إختبار الفرضيات	Test of Hypothesis
إختبارات إحصائية	Statistics Tests
أخطاء	Errors
أداة	Tool
أدنى	Lower
إرتباط أو تلازم	Correlation
إرتباط بين الصفات	Correlation between Attributes
أرقام قياسية	Index numbers
أساليب	Approaches
إستبانة	Questionnaire

الملحق A

المصطلحات العلمية

Sampling Frame	إطار المعاينة
Upper	أعلى
Bar Charts	أعمدة بيانية
Simple Bar Chart	أعمدة بيانية بسيطة
Clustered or Stacked Bar Chart	أعمدة بيانية مركبة
Bars	أعمدة مستطيلة
Minimize	أقل مايمكن
Skew ness	إلتواء
Scatter	إنتشار
Simple Linear Regression	إنحدار خطي بسيط
Multiple Linear Regression	إنحدار خطي متعدد
Deviation	إنحراف
Quartile Deviation	إنحراف ربيعي
Mean Deviation	إنحراف متوسط
Standard Deviation	إنحراف معياري
Weights	أوزان
Prior	أولية

الملحق A

المصطلحات العلمية

Residuals	بواقي
Data	بيانات
Statistical data	بيانات إحصائية
Nominal data	بيانات إسمية
Ordinal data	بيانات ترتيبية
Qualitative data	بيانات صفات نوعية
Not- tabulated Data	بيانات غير مبوبة
Quantitative data or Scale	بيانات كمية
Tabulated Data	بيانات مبوبة
Continuous data	بيانات متصلة
Discrete data	بيانات متقطعة
Exact	تامة
Variance	تباين
Covariance	تباين مشترك
Data Tabulation and Presentation	تبويب وعرض البيانات
Homogenous	تجانس
Analysis	تحليل

الملحق A

المصطلحات العلمية

Regression Analysis	تحليل الإنحدار
Analysis Data	تحليل البيانات
Biased	تحيز
Increasing	تصاعدي
Classical	تصنيف
Due to	تعزى إلى
Changes	تغيرات
Interpreting	تفسير
Kurtosis	تفلطح (تفرطح)
Intercept	تقاطع خط الإنحدار مع المحور العمودي
Estimation	تقدير
Interval Estimation	تقدير فترة
Point Estimation	تقدير نقطي
Class Frequency	تكرار الفئة
Decreasing	تنازلي
Prediction	تنبؤ
Frequency Distribution	توزيع تكراري

الملحق A

المصطلحات العلمية

Cumulative Frequency Distribution	توزيع تكراري متجمع
Proportionate Frequency Distribution	توزيع تكراري نسبي
Normal Distribution	توزيع متماثل (توزيع طبيعي)
Expectation	توقع
Constant	ثابت
Confidence	ثقة
3-Dimintion	ثلاثي الأبعاد
Tables	جداول
Frequency Table	جدول تكراري
Contingency Table	جدول توافق
Root	جذر
Size	حجم
Lower Limit of a Class	حد الأدنى للفئة
Upper Limit of a Class	حد الأعلى للفئة
Limits	حدود
Goodness of Fit	حسن المطابقة (ملائمة النموذج)
Properties	خصائص

الملحق A

المصطلحات العلمية

Line Chart	خط بياني
Standard Error	خطأ معياري
Random Error	خطأ عشوائي
Linear	خطي
Pie Chart	دائرة بيانية
Function	دالة
Demography	دراسات سكانية
Degree of Freedom	درجة الحرية
accuracy	دقة
Lower Quartile	ربيع أدنى
Upper Quartile	ربيع أعلى
Quartiles	ربيعات
Rank	رتبة
Graphical	رسوم بيانية
Reject	رفض
Time	زمن
Even	زوجي

الملحق A

المصطلحات العلمية

Negative	سالبة
Series	سلسلة
Behavior	سلوك
Outlier	شاذه
Chance	صدفة
Row	صف
Formula	صيغة
Random Samples	طرائق عشوائية
Nonrandom Samples	طرائق غير عشوائية
Statistical Method	طريقة إحصائية
Ordinary Least Squares Method	طريقة المربعات الصغرى الإعتيادية
Length of Class	طول الفئة
Number of Classes	عدد الفئات
Graphical Presentation	عرض بياني
Moment	عزوم
Relationship	علاقة
Process	عملية

الملحق A

المصطلحات العلمية

Column	عمود
Sample	عينة
Accidental Sample	عينة بالمصادفة
Quota Sample	عينة حصصية
Multi-stage Random Sample	عينة ذات مراحل متعددة
Simple Random Sample	عينة عشوائية بسيطة
Stratified Random Sample	عينة عشوائية طبقية
Purposive Sample	عينة عمدية
Non-Linear	غير خطي
Un measurable	غير قابلة للقياس
Unbiased	غير متحيز
Class	فئة
Odd	فردي
Measurable	قابلة للقياس
Accept	قبول
Blocks	قطاعات
Laws	قوانين

الملحق A

المصطلحات العلمية

Scales	قياسات
Standard value	قيمة معيارية
Expected Value	قيمة مقدرة
Quality Control Charts	لوحات السيطرة النوعية
Logarithm	لوغارتم
Indexes	مؤشرات
Extreme	متطرفة
Multivariate	متعدد المتغيرات
Variable	متغير
Dependent Variable	متغير تابع أو معتمد
Random Variable	متغير عشوائي
Independent Variable	متغير مستقل
Univariate	متغير واحد
Bivariate	متغيرين
Symmetric	متماثل
Mean	متوسط (وسط حسابي)
Optimum	مثالية

الملحق A

المصطلحات العلمية

Population	مجتمع
Statistical Population	مجتمع إحصائي
Infinite Population	مجتمع غير محدود
Finite Population	مجتمع محدود
Summation	مجموع
Group	مجموعة
Deterministic	محددة
Horizontal Axis	محور أفقي
Vertical Axis	محور عمودي
Histogram	مدرج تكراري
Range	مدى
Total Range	مدى كلي
Stages	مراحل
Square	مربع
Center of a Class	مركز الفئة
Double	مزدوج
Level	مستوى

الملحق A

المصطلحات العلمية

Confidence level	مستوى الثقة
Observation	مشاهدات
Real Observations	مشاهدات حقيقية
Frequency Polygon	مضلع تكراري
Absolute	مطلقة
Equations	معادلات
Parameters	معالم أو معلمات
Coefficient	معامل
Coefficient of Variation	معامل الإختلاف
Partial Correlation Coefficient	معامل الإرتباط الجزئي
Simple Linear Correlation Coefficient	معامل الإرتباط الخطي البسيط
Multiple Correlation Coefficient	معامل الإرتباط المتعدد
Coefficient of Association	معامل الإقتران
Regression Coefficient	معامل الإنحدار
Coefficient of Determination	معامل التحديد
Coefficient of Dispersion	معامل التشتت
Coefficient of Contingency	معامل التوافق

الملحق A المصطلحات العلمية	
معاينة	Sampling
معدل	Average
معكوس	Inverse
معلومات	Information
مفقودة	Missing
مقاييس التشتت	Measures of Variation
مقاييس النزعة المركزية	Measures of Central Tendency
مقدر	Estimator
منحني	Curve
منحنى تكراري	Frequency Curve
منحنى تكراري متجمع	Cumulative Frequency Curve
منوال	Mode
موجبة	Positive
ميل	Slop
نسب	Proportions
نظرية إحصائية	Statistical Theory
نمو	Growth

الملحق A

المصطلحات العلمية

Model	نموذج
Sampling Units	وحدات المعاينة
Quadratic Mean	وسط تربيعي
Harmonic Mean	وسط توافقي
Weighted Mean	وسط حسابي موزون (المرجح)
Geometric Mean	وسط هندسي
Median	وسيط

الملحق B
الصيغ الإحصائية

المؤشر الإحصائي	الصيغة الرياضية
حجم العينة على أساس المتوسط	$n = \frac{z^2\sigma^2}{e^2}$
حجم العينة على أساس النسب	$n = \frac{z^2 P(1-P)}{e^2}$
المدى الكلي	$T.R = x_L - x_S + 1$
عدد الفئات- طريقة يول	$m = (2.5)\cdot\sqrt[4]{n} = (2.5)n^{1/4}$
عدد الفئات-طريقة سترجس	$m = 1 + (3.322)\log_{10}(n)$
طول الفئة	$L = \frac{T.R}{m}$
مركز الفئة	$x = \frac{L.L + U.L}{2}$
التكرارات النسبية	$f_i^* = \frac{f_i}{n}.100$

الملحق B
الصيغ الإحصائية

التوزيع التكراري المتجمع الصاعد	$F_1 = f_1$ $F_2 = f_1 + f_2$ $\vdots$ $F_m = f_1 + f_2 + \cdots + f_m$
توزيع تكراري صاعد نسبي	$F_i^* = \frac{F_i}{n}.100$
التوزيع التكراري المتجمع النازل	$F_1' = n$ $F_2' = n - f_1$ $\vdots$ $F_m' = n - f_1 - f_2 - \cdots - f_{m-1} = f_m$
توزيع تكراري نازل نسبي	$F_i'^* = \frac{F_i'}{n}.100$
تكرار معدل (عدم تساوي أطوال الفئات)	$f_i^* = \frac{f_i}{L_i} \quad ; \quad i = 1,2,\cdots,m$
زاوية القطاع للدائرة البيانية	$\theta_i = \frac{A_i}{T} \times 360^\circ$
الوسط الحسابي للبيانات غير المبوبة	$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

الملحق B الصيغ الإحصائية	
معدل المجتمع	$\mu = \dfrac{\sum_{i=1}^{N} x_i}{N}$
الوسط الحسابي للبيانات المبوبة	$\overline{x} = \dfrac{\sum_{i=1}^{m} f_i x_i}{\sum_{i=1}^{m} f_i}$
الوسط الحسابي الموزون للبيانات غير المبوبة	$\overline{x}_w = \dfrac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$
الوسط الحسابي الموزون للبيانات المبوبة	$\overline{x}_w = \dfrac{\sum_{i=1}^{m} w_i f_i x_i}{\sum_{i=1}^{m} w_i f_i}$
الوسط التوافقي للبيانات غير المبوبة	$H = \dfrac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$
الوسط التوافقي للبيانات المبوبة	$H = \dfrac{\sum_{i=1}^{m} f_i}{\sum_{i=1}^{m} \frac{f_i}{x_i}}$

الملحق B الصيغ الإحصائية	
الوسط التربيعي للبيانات غير المبوبة	$$Q = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$$
الوسط التربيعي للبيانات المبوبة	$$Q = \sqrt{\frac{\sum_{i=1}^{m} f_i x_i^2}{\sum_{i=1}^{m} f_i}}$$
الوسط الهندسي للبيانات غير المبوبة-1	$$G = \sqrt[n]{\prod_{i=1}^{n} x_i} = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$$
الوسط الهندسي للبيانات غير المبوبة-2	$$G = anti - Log_{10}\left(\frac{1}{n}\sum_{i=1}^{n} Log_{10} x_i\right)$$
الوسط الهندسي للبيانات المبوبة-1	$$G = \sqrt[\sum_{i=1}^{m} f_i]{\prod_{i=1}^{m} x_i^{f_i}} = \left(\prod_{i=1}^{m} x_i^{f_i}\right)^{\frac{1}{n}}$$
الوسط الهندسي للبيانات المبوبة-2	$$= anti - \log_{10}\left(\frac{1}{n}\sum_{i=1}^{m} f_i \log_{10} x_i\right)$$
الوسيط للبيانات المبوبة من خلال التوزيع التكراري الصاعد	$$Me = L_i + \left[\frac{n/2 - F_i}{f_i}\right] \cdot w$$

الملحق B

الصيغ الإحصائية

الوسيط للبيانات المبوبة من خلال التوزيع التكراري النازل	$$Me = L_i + \left[\frac{F_i' - n/2}{f_i} \right] \cdot w$$
المدى للبيانات غير المبوبة	$$R = x_l - x_s$$
المدى للبيانات المبوبة	حساب الفرق مابين مركز الفئة العليا والدنيا
الإنحراف المتوسط للبيانات غير المبوبة	$$M.D = \frac{\sum_{i=1}^{n} \lvert x_i - \bar{x} \rvert}{n}$$
الإنحراف المتوسط للبيانات المبوبة	$$M.D = \frac{\sum_{i=1}^{m} f_i \lvert x_i - \bar{x} \rvert}{\sum_{i=1}^{m} f_i}$$
الإنحراف الربيعي للبيانات غير المبوبة	$$Q \cdot D = \frac{Q_u - Q_l}{2}$$
الإنحراف المعياري للبيانات غير المبوبة-1	$$S = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

الملحق B

الصيغ الإحصائية

$$S = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}}$$	الإنحراف المعياري للبيانات غير المبوبة-2
$$S = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1}}$$	الإنحراف المعياري للبيانات غير المبوبة-3
$$S = \sqrt{\frac{\sum_{i=1}^{m} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{m} f_i - 1}}$$	الإنحراف المعياري للبيانات المبوبة
$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$	التباين للبيانات غير المبوبة
$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N-1}$$	تباين المجتمع
$$S^2 = \frac{\sum_{i=1}^{m} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{m} f_i - 1}$$	التباين للبيانات المبوبة

الملحق B الصيغ الإحصائية	
$C\cdot V = \frac{S}{\overline{x}}\cdot 100$	معامل الإختلاف
$r_{xy} = \frac{\operatorname{cov}(x, y)}{S_x \,.\, S_y} = \frac{S_{xy}}{S_x \,.\, S_y}$	معامل الإرتباط الخطي البسيط
$\operatorname{cov}(x, y) = S_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n-1}$	التباين المشترك
$r_{xy} = 1 - \frac{6\cdot\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$	معامل إرتباط الرتب لسبيرمان (في حالة عدم وجود تكرار)
$r_{xy} = 1 - \frac{6\cdot\left(\sum_{i=1}^{n} d_i^2 + k\right)}{n(n^2-1)}$	معامل إرتباط الرتب لسبيرمان (في حالة وجود تكرار)
$r_{12\,.\,3} = \frac{r_{12} - r_{13}\,r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$	معامل الإرتباط الجزئي لثلاث متغيرات
$r_{12\,.\,34} = \frac{r_{12\,.\,3} - r_{14\,.\,3}\,r_{24\,.\,3}}{\sqrt{(1 - r_{14\,.\,3}^2)(1 - r_{24\,.\,3}^2)}}$	معامل الإرتباط الجزئي لأربعة متغيرات
$r_{1\,.\,23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2\,r_{12}\cdot r_{13}\cdot r_{23}}{1 - r_{23}^2}}$	معامل الإرتباط المتعدد

الملحق B الصيغ الإحصائية	
لثلاث متغيرات-1	
معامل الإرتباط المتعدد لثلاث متغيرات-2	$r_{1.23} = \sqrt{1-\left(1-r_{12}^{2}\right)\left(1-r_{13.2}^{2}\right)}$
معامل الإرتباط المتعدد لأربعة متغيرات	$_{4} = \sqrt{1-\left(1-r_{12}^{2}\right)\left(1-r_{13.2}^{2}\right)\left(1-r_{14.23}^{2}\right)}$
معامل الإرتباط المتعدد إلى k من المتغيرات	$- r_{12}^{2}\big)\left(1-r_{13.2}^{2}\right)\left(1-r_{14.23}^{2}\right)\cdots\left(1-r_{1k.23\ldots k-1}^{2}\right)$
معامل التوافق	$C = \sqrt{\frac{r-1}{r}}$
معامل الإقتران في حالة مستويين	$C.A_2 = \frac{f_{11}\cdot f_{22} - f_{12}\cdot f_{21}}{f_{11}\cdot f_{22} + f_{12}\cdot f_{21}}$
معامل الإقتران في حالة مستويين أو أكثر	$C.A_G = \frac{\sum_{i=1}^{m} f'_{ij} + \sum_{j=1}^{k} f''_{ij} - T'_{.i} - T'_{j.}}{2n - \left(T'_{.i} + T'_{j.}\right)}$
نموذج الإنحدار الخطي البسيط	$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$
معامل الإنحدار الخطي البسيط	$\hat{\beta}_1 = \frac{S_{xy}}{S_x^2} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\cdot\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2}$

الملحق B

الصيغ الإحصائية

نقطة تقاطع خط الإنحدار مع المحور العمودي	$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
البواقي	$e_i = y_i - \hat{y}_i$
علاقة بين معامل الإنحدار $\hat{\beta}_1$ ومعامل الإرتباط البسيط	$\hat{\beta}_1 = \frac{S_y}{S_x} \cdot r_{xy}$
معامل التحديد-1	$R^2 = (r_{xy})^2$
معامل التحديد-2	$R^2 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$
معامل التحديد-3	$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$
الخطأ المعياري	$S_{y/x} = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-2}}$
نموذج الإنحدار المتعدد ذات متغيرين مستقلين	$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}$

الملحق B

الصيغ الإحصائية

$$\frac{{}_{1}Y_{i}\Big)\Big(\sum_{i=1}^{n} X_{i2}^{2}\Big)-\Big(\sum_{i=1}^{n} X_{i2}Y_{i}\Big)\Big(\sum_{i=1}^{n} X_{i1}X_{i2}\Big)}{\Big(\sum_{i=1}^{n} X_{i1}^{2}\Big)\Big(\sum_{i=1}^{n} X_{i2}^{2}\Big)-\Big(\sum_{i=1}^{n} X_{i1}X_{i2}\Big)^{2}}$$	معامل الإنحدار إلى x_1
$$\frac{{}_{i2}Y_{i}\Big)\Big(\sum_{i=1}^{n} X_{i1}^{2}\Big)-\Big(\sum_{i=1}^{n} X_{i1}Y_{i}\Big)\Big(\sum_{i=1}^{n} X_{i1}X_{i2}\Big)}{\Big(\sum_{i=1}^{n} X_{i1}^{2}\Big)\Big(\sum_{i=1}^{n} X_{i2}^{2}\Big)-\Big(\sum_{i=1}^{n} X_{i1}X_{i2}\Big)^{2}}$$	معامل الإنحدار إلى x_2
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}_1 - \hat{\beta}_2\bar{x}_2$$	نقطة تقاطع مستوى الإنحدار مع المحور العمودي
$$R^{2}_{y\cdot x_1x_2} = \left(r_{y\cdot x_1x_2}\right)^2$$	معامل التحديد للإنحدار الخطي المتعدد (ذات متغيرين)-1
$$R^{2}_{y\cdot x_1x_2} = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} Y_i^2}$$	معامل التحديد للإنحدار الخطي المتعدد (ذات متغيرين)-2

الملحق B الصيغ الإحصائية	
$$R^2_{y\bullet x_1x_2} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{n} Y_i^2}$$	معامل التحديد للإنحدار الخطي المتعدد (ذات متغيرين)-3
$$R^2_{y\bullet x_1x_2} = \frac{\hat{\beta}_1 \cdot \sum_{i=1}^{n} X_{i1}Y_i + \hat{\beta}_2 \cdot \sum_{i=1}^{n} X_{i2}Y_i}{\sum_{i=1}^{n} Y_i^2}$$	معامل التحديد للإنحدار الخطي المتعدد (ذات متغيرين)-4
$$S_{y/x_1x_2} = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-3}}$$	الخطأ المعياري للإنحدار المتعدد (ذات متغيرين)

المصادر

أولاً المصادر العربية:

1. الراوي، د.خاشع، (1980)، **المدخل إلى الإحصاء**، جامعة الموصل.

2.المشهداني، د.محمود حسن وأمير حنا هرمز، (1989)، **الإحصاء**، جامعة بغداد-بيت الحكمة.

3. الناصر، د.عبد المجيد حمزة الناصر ود.عصرية ردام المرزوك، (1989)، **العينات**، جامعة بغداد-بيت الحكمة.

4. الناصر، فوزي عبد الرزاق، (1985)، مبادئ الإحصاء الحديث (مترجم)، الجامعة المستنصرية، مطبعة جامعة الموصل.

5. الهيتي، د.صلاح الدين حسين، (2006)، **الأساليب الإحصائية في العلوم الإدارية-تطبيقات باستخدام** SPSS، جامعة مؤتة.

6. جودة، د.محفوظ،(2008)، **التحليل الإحصائي الأساسي باستخدام** SPSS، جامعة العلوم التطبيقية-مطبعة وائل.

ثانياً المصادر الأجنبية:

Series	References
7	Agresti, A. (1990). Categorical Data Analysis. New York: John Wiley and Sons.
8	Ashford, J.R. (1959). An approach to an analysis of data for semiquantal responses in biological assay. Biometrics.
9	Azzalini,A.(1996)، Statistical Inference Based on the Likelihood London: Chapman and Hall.
10	Barndorff-Nielsen, O.E. (1988). Parametric Statistical Models and Likelihood. New York: Springer.
11	Barndorff-Nielsen, O.E. and Cox، D.R. (1989). Asymptotic Techniques for Use in Statistics. London: Chapman and Hall.
12	Birnbaum, A. (1962). On the foundations of statistical inference. Journal of the American Statistical Association، 57.
13	Brookes, B.C. and Dick، W.F.L. (1951). Introduction to Statistical Method. London: Heinemann.
14	Christensen, R. (1990). Log-Linear Models. Berlin: Springer.
15	Collet, D. (1991). Modelling Binary Data. London: Chapman and Hall.
16	Cook, R.D. (1977). Detection of influential observations in linear regression. Techno metrics، 19.
17	Cook, R.D. and Weisberg, S. (1982). Residuals and Influence in Regression. London: Chapman and Hall.

18 Cox, D.R. (1972). Regression models and life tables (with discussion). Journal of the Royal Statistical Society B، 34.

19 Cox, D.R. (1990). Role of models in statistical analysis. Statistical Science، 5.

20 Cox, D.R. and Hinkley، D.V. (1974). Theoretical Statistics. London: Chapman and Hall.

21 Dalgaard, P. (2002). Introductory Statistics with R. New York: Springer.

22 Davison,A.C. (2003). Statistical Models. Cambridge: Cambridge University Press.

23 Dobson, A.J. (1990). An Introduction to Generalized Linear Models. London: Chapman and Hall.

24 Draper,N.R.and Smith، H. (1981). Applied Regression Analysis. 2nd ed. New York: John Wiley and Sons.

25 Faraway, J.J. (2005). Linear Models with R. Boca Raton, FL: Chapman and Hall/CRC.

26 Fisher,R.A.(1956).Statistical Methods and Scientific Inference. Edinburgh: Oliver and Boyd.

27 Fraser, D.A.S. (1979). Inference and Linear Models. New York: McGraw Hill.

28 Haberman, S.J. (1978). Analysis of Qualitative Data. Volume 1. Introductory Topics. San Diego، CA: Academic Press.

29 Knight,K.(2000). Mathematical Statistics. New York: Chapman and Hall.

30 Lee,E.T.(1992).Statistical Methods for Survival Data Analysis. New York: John Wiley and Sons.

31 Linhart, H. and Zucchini، W. (1986). Model Selection. New York: John Wiley and Sons.

32 Lindsey, J.K. (1995a). Modelling Frequency and Count Data. Oxford: Clarendon Press.

33 Lindsey, J.K. (1995b). Introductory Statistics: The Modelling Approach. Oxford: Oxford University Press.

34 Lindsey, J.K. (1996). Parametric Statistical Inference. Oxford: Clarendon Press.

Series References

35 Lindsey, J.K. (1997). Applying Generalized Linear Models. New York: Springer.

36 Miller, R.G. (1981). Survival Analysis، New York: John Wiley and Sons.

37 Searle, S.R. (1971). Linear Models. New York: John Wiley and Sons.

38 Seber, G.A.F. (1977). Linear Regression Analysis. New York: John Wiley and Sons.

39 Silvey, S.D. (1980). Statistical Inference. London: Chapman and Hall.

40 Steel,R.G.D. and Torrie، J.H. (1980). Principles and Procedures of Statistics. New York: McGraw Hill.

41 Stigler,S.M. (1986). The History of Statistics: The Measurement of Uncertainty Before 1900. Cambridge, MA: Belknap Press.

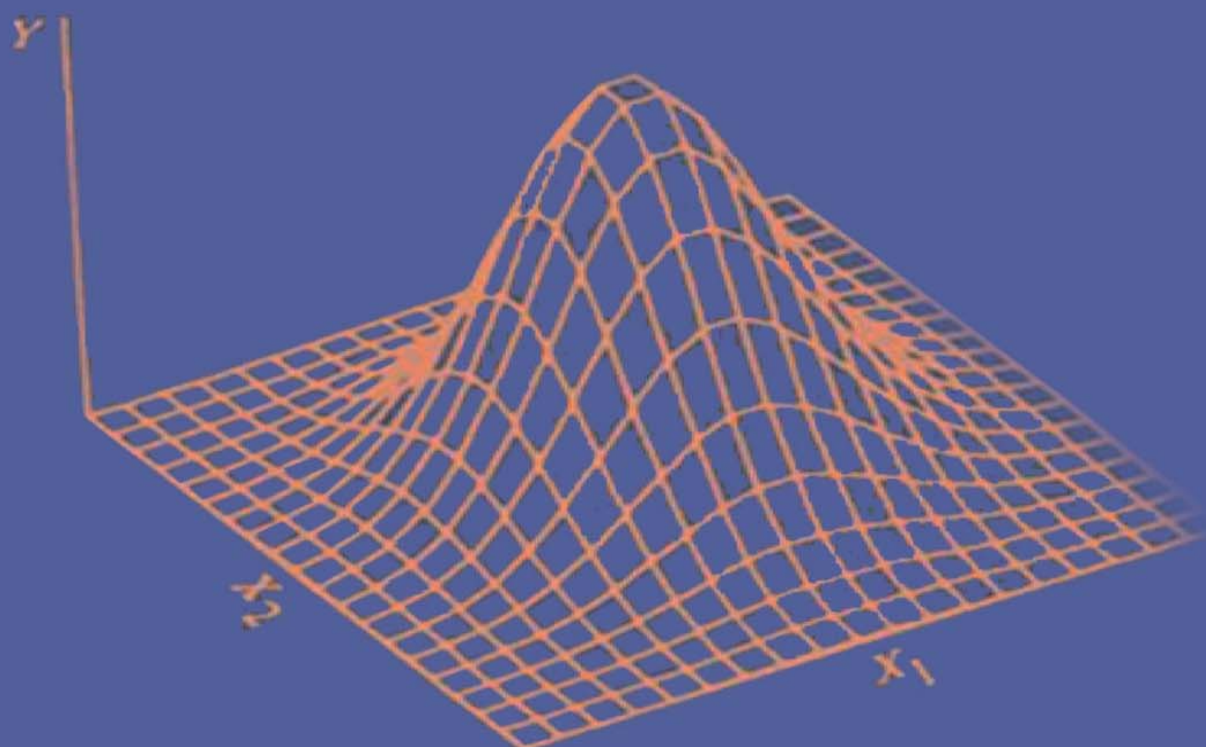
42 Stirzaker, D.R. (1994). Elementary Probability. Cambridge: Cambridge University Press.

43 Uusipaikka, E.I. (2006). Statistical Inference Package SIP. http://www.wolfram.com/products/applications/sip/.

44 Weisberg, S. (1985). Applied Linear Regression. 2nd ed. New York: John Wiley and Sons.

45 Welsh, A.H. (1996). Aspects of Statistical Inference. New York: John Wiley and Sons.

46 Wetherill, G.B. (1986). Regression Analysis with Applications. London: Chapman and Hall.







دار غيداء للنشر والتوزيع

تلاع العلي - شارع الملكة رانيا العبدالله
مجمع العساف التجاري - الطابق الأول

تلفاكس : 5353402 6 962+
خلـــوي : 95667143 7 962+

ص.ب : 520946 عمان 11152 الأردن
E-mail: darghidaa@gmail.com